Studying Gestures: The Iconic Roots of Human Communication Systems

B002562

MSc Evolution of Language and Cognition

The University of Edinburgh

2011

Table of Contents

ACKNOWLEDGEMENTS

I am sincerely indebted and grateful to my supervisors, Kenny Smith, Monica Tamariz, and Erin Brown for the support and guidance they showed me throughout my dissertation writing. I am sure it would have not been possible without their help. I would also like to thank Nic Fay for his useful advice and comments on my experiment. Special thanks to my flatmates and good friends, Constantinos Christofi and Andreas Iacovou, for their support throughout the year and their help in running my experiments. Most importantly I would like to thank my girlfriend Chrystalla Kyriakou (for putting up with me this long year), my family, and especially my mum for teaching me how to live a decent life and giving me absolutely everything without ever asking anything in return. LOVE YOU MUM

Studying Gestures: The Iconic Roots of Human Communication Systems

Abstract

This paper tests the hypothesis that the iconicity inherent in human gestures can be a key element in the creation and evolution of communication systems. An interactive experiment based on playing charades was conducted modelling a situation where a multimodal (gesture + sound) and semiotically mixed (non-arbitrary + arbitrary) communication system evolves through social collaboration. The experimental design was based on Garrod and colleagues' (2007) modelling of iconic graphical signals' transformation into symbolic ones through interactive communication experiments. The reported experiment, however, diversified from the original in certain ways; (a) graphical representations were replaced by gestural representations (i.e. pantomime); (b) a vocalization component became mandatory in one of the experimental conditions; and (c) feedback exchange between participants was not allowed. Comparing the evolution of gestural communication across two separate conditions ('Gesture-only', 'Gesture-and-sound'), the experiment reveals that in the absence of feedback, simplification (i.e. reduction of forms) and refinement of signs is still attainable; albeit with significant differences across the two conditions. It is possible that pantomimic gestures *can* serve as a bootstrapping device between meanings and arbitrary sounds – even in the absence of feedback – if the latter are repeatedly associated with iconic gestural signs in the context of reciprocal pair-wise interactions. Although gestures in the experiment lose some of their initial iconic features due to their simplification and refinement, it cannot be argued that they have transformed into symbols – gestures can remain iconic and at the same time be shared between people who have never interacted before.

1. Introduction

Gesture is an elemental component of human discourse (de Jorio 1832; McNeill 1992, 2005). When the vocal-auditory channel is unable to undertake the task of human communication for whatever reason (e.g. silence oath in Australian Aborigines, see Kendon 1988; deafness, see Goldin-Meadow & Mylander 1990; working in a noisy factory, see Barakat 1969, Brun 1969), gesture is often employed by humans to fill in this void. Is it likely that gestures were a decisive factor in the initial stages of language evolution? In other words, is there a possibility that language evolved firstly as gesture and then transferred to the oral modality – instead of speech dominating language from the very beginning?

The purpose of this paper will be to test the hypothesis that the iconicity inherent in human gestures can be a key element in the creation and evolution of communication systems. In doing so we will examine the conditions under which a multimodal (gesture + sound) and semiotically mixed (non-arbitrary + arbitrary) communication system using both gestures (pantomime) and arbitrary sounds can emerge as a result of repeated interaction. Reciprocal pair-wise interactions are expected to result in the progressive convergence of signs within pairs; local interactions produce accurate local sign systems that are shared within the interacting dyads (see Garrod et al. 2007, Fay et. al. 2008). Examining the iconic roots of communication systems, an experiment was conducted dealing with the evolution of a gestural communication system (i.e. based on pantomime) and its usefulness for establishing a communication system in the vocal-auditory channel. It will be argued that iconicity is highly important for establishing communication in the first place. Additionally, once communication is established, interaction – involving imitation and some kind of indirect feedback – consolidates faster and more efficient communication.

The central experimental hypothesis of this dissertation evolves around two separate but strongly interconnected axes: (a) the creation of a shared communication system (a lexicon) based on gestures through a "grounding process" (Clark & Brennan, 1991); (b) this gestural system can be used to lend meaning to previously meaningless sounds. The experimental design was based on Garrod and colleagues' (2007) modelling of iconic graphical signals' transformation into symbolic ones through interactive communication

experiments. My experiment entailed, however, some small but highly significant modifications; (a) graphical representations were replaced by gestural representations (pantomime); (b) a vocalization component became mandatory in one of the experimental conditions; and (c) feedback exchange between participants was not allowed.

One of the basic premises of Garrod and colleagues' (2007) study is that icons can transform into symbols, "supported by an interactive grounding process" (p.961). It will be argued that although gestures in my experiment lose some of their initial iconic features due to their simplification and refinement, they do not transform in any way into symbols. Gestures can thus remain iconic for the people who create them (i.e. hold a non-arbitrary relation to their referents) and at the same time be shared between people who have never interacted before. The discussed experiment provides an alternative situation in which symbolic (arbitrary) communication can be established; symbols *can* arise in a situation involving a multimodal (gesture & sound) and semiotically mixed (non-arbitrary & arbitrary) communication system. Despite being unnecessary for successful communication, formerly meaningless sounds can come to carry meaning in an arbitrary way simply by being there and repeatedly being associated with meaningful iconic gestures.

After a brief introduction in which we clarify the use of the term 'gesture' in this paper (Section 2), we will present how several authors proposing putting forward various gestural hypotheses of language origins deal with the issue of iconicity in the context of language evolution(Section 3). In Section 4, the Garrod et al. (2007) study, on which the discussed experiment is based, will be briefly presented. The design of the experiment will be presented in Section 5, followed by the obtained results (Section 6) and a detailed discussion of their implications (Section 7). The last Section (9) comprises of the paper's concluding remarks along with some proposals for future applications of the discussed experimental design.

2. Defining gestures

Even though language tends to be regarded as a system of solely symbolic, conventional and/or arbitrary signs, we should not ignore the fact that it also encompasses elements based greatly on iconicity of non-linguistic reality (Landsberg 1980). Apart from the iconicity we encounter in speech, many linguists have started to consider the possibility that the gestures accompanying speech are also firmly connected to the meaning we intend to convey (see Goldin-Meadow 2005; Kendon 1997; McNeill 1992, 2005). The first thing we need to clarify is what we actually mean by the term 'icon' and what is its status in the realm of semiotics, i.e. the study of signalling systems. In keeping with Peirce (1931-1958), signs can be divided into two categories; 'indices' and 'signals', with the latter further divided into 'symbols' and 'icons' (ibid.). 'Indices' are connected causally to their referents; 'symbols' have an arbitrary and/or conventional relationship with the objects they refer to; and 'icons' bear a salient perceptual/ structural similarity to their referents (Garrod et al. 2007, p. 964). What links indices and icons is that both have a non-arbitrary relationship with their referents[1]. Garrod and colleagues (2007) claim that it is rather impossible to define a firm line dividing symbols from icons; the distinction between them is fairly gradient. It appears that iconicity is more prominent in cases where the information is organized around the structure of the sign, whereas in cases where the information is based on the agent's knowledge of the sign's prior usage, symbolism is to be expected (ibid.).

Before embarking on experimental work, it is vital to have a clear view of what we actually mean by gestures when we refer to them. In spite of the fact that gestures carry within them a high degree of iconicity this does not mean that dissociation from it is impossible (e.g. sign languages). This paper adapts McNeill's (1992, 2005) criteria for distinguishing different types of gestures based on Kendon's (1981) ideas. The division is based on the different ways these gestures convey meaning:

(a) Gesticulations: (iconics, metaphorics, deictics, beats): Gestures accompanying and coordinating in time and space with language (either spoken or signed)

[1] From now on the term 'non-arbitrariness' will be used to refer to both iconicity and indexicality.

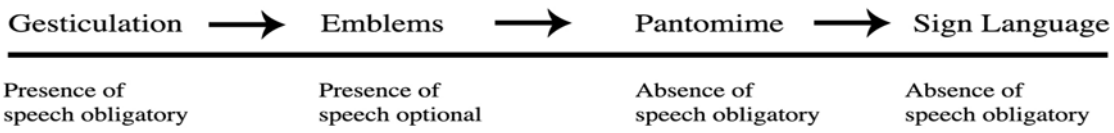
(b) Pantomime: Gestures produced to convey some kind of meaning – basically through trying to represent iconically the intended referent

(c) Emblems: Gestures that have become conventionalized and hold an arbitrary relationship with the meaning they signify

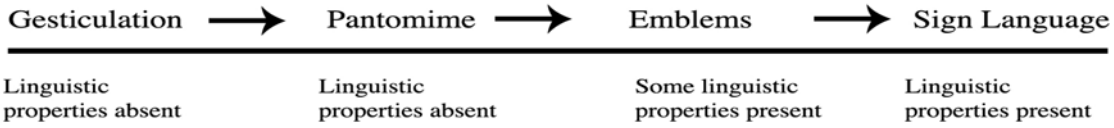
(d) Signs: Gestures that are part of a linguistic system such as sign languages and which have gradually become conventionalized and arbitrary – although the former does not presuppose the latter

McNeill arranges these four types of gestures on four separate continua based on their relationship to different variables (Figures adopted from Parrill 2008):

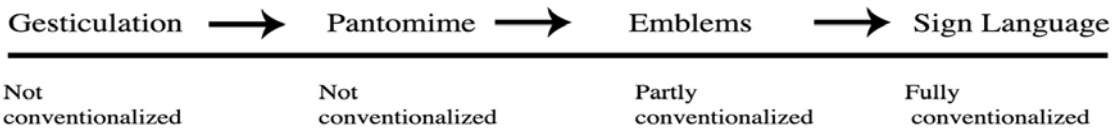
(a) Connection to speech



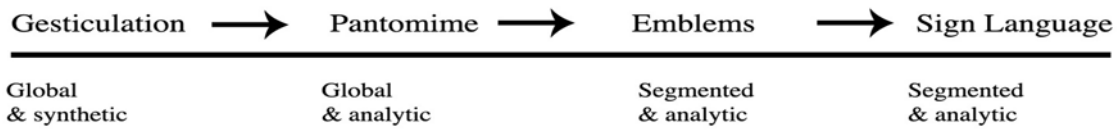
(b) Similarity to a linguistic system



(c) Degree of conventionalization



(d) Semiotic distinctions between the four gesture types



Gesture is inherent in human communication, something that is manifested from the very first stages of language development in human ontogeny. Studying the transition from one- to two-word stage in 10 normally developing children, Iverson and Goldin-Meadow (2005) discovered that the first items to enter the children's spoken vocabulary were those firstly produced by children gesturally(e.g. initially most children used a pointing gesture to refer to a 'ball' before moving on to produce its spoken form). Moreover, children who used gesture-plus-word supplementary combinations (i.e. combinations with no redundant elements) to communicate a simple proposition (e.g. pointing to a bird and uttering 'nap' to communicate that 'the bird is sleeping') were the first to generate the same combination exclusively in speech (e.g. uttering 'bird nap') (ibid.). In another study, Özçalişkan and Goldin-Meadow (2005) noticed that the supplementary gestures of their infant participants rose significantly in numbers as they moved from 14 to 22 months of age and consistently foreshadowed changes in a child's verbal output. For example, 18-month-olds were able to generate a gesture-plus-speech combination for an argument-predicate structure (e.g. uttering 'you' and pantomiming 'hit'); this ability was expressed exclusively in speech only a few months later (ibid.)[2]. Hence, gesture develops and progresses in tandem with speech – children's earliest communication being fundamentally multimodal. These observations reveal the possibility of the existence of a general 'representational' ability underlying the emergence of gestures (pointing and pantomimes) and conventional language (Brown in prep. a)[3].

In modern humans, gesticulations (speech-synchronized gestures) are an essential part of language and not solely an accompaniment of conventional linguistic forms (McNeill 2005, McNeill et al. 2008). Gesticulations are "semantically and pragmatically co-expressive

[2] The children were followed longitudinally between the ages of 10 to 24 months.
[3] The same argument and examples were used in the final paper I submitted in the 'Biolinguistics' course (Semester 2).

with speech, (but) not redundant (McNeill et al. 2008, p.2). Hence, human discourse is constituted from two different, but synchronous, semiotic modes – one instantiated in speech-synchronized gestures and one instantiated in speech (ibid.). The contemporaneousness of the two reveals the dual nature of human communication, what McNeill (2005), calls the 'imagery-language' dialectic. Gesticulations are uninterpretable unless they surface with speech (McNeill 2005). Having that said, this non-interpretability does not disqualify them from expressing an imagistic representation of a particular semantic content (Brown in prep. a), and this representation does not have to be different from the one expressed by pantomimic bodily actions. If I want, for example, to pantomime kicking a football, I will form my hands in a fashion expressing my holding of an imaginary ball[4]. I will then lower down my trunk, and finally enact the kick by moving my foot up and forward. If I would describe this event verbally saying something like, "I took the ball and kicked it with my foot", my gesticulation would have probably involved solely my hands, using them to show the shape and size of the ball, but also pictorially representing the act of kicking by abruptly moving my hand up and forward (Note in this case that the shape and size of the ball are not expressed in speech but solely in gesture, revealing the co-expressiveness and synchrony of the two. The abruptness of my movement may also communicate the power with which I kicked the ball as well as its trajectory, again not surfacing in my utterance). The intended meaning in both cases was probably triggered by the same representation of what is registered in my mind as the concept of kicking a football. Brown (in prep. a) argues that the dissimilarities in surface structure are, at least partly, the outcome of the absence (or presence) of an accompanying linguistic context. The resemblance between gesticulation and pantomime lies on the fact that they are both based on iconicity and idiosyncrasy, and they both involve some kind of re-enacting an intended meaning in our head. In the presence of speech (or signs) pantomime tends to disappear, while gesticulation is still firmly connected to it.

McNeill and colleagues (2008) argue that pantomime and gesticulation are fundamentally different (the two having followed diverging evolutionary trajectories in the course of human evolution) due to the dissimilar temporal relationship the two share with speech. Hence, they see pantomime as having an 'alternating' (as opposed to 'synchronous')

[4] This example is taken from the final paper I submitted in the 'Biolinguistics' course (Semester 2)

temporal relationship with speech (p.10) – "pantomime, by definition, does not accompany speech" (McNeill 2005, p.7). However, this might be a result of a human convention set on pantomime; we usually expect people not to speak when they perform pantomimic actions, but this does not mean, for example, that I cannot mimic a flying bird, and simultaneously start jumping around screaming "I'm a bird, I'm a bird"[5]. Moreover, in a Fay and colleagues' (2010) study, participants being asked to use non-arbitrary vocalizations in order to communicate, still pantomimed the meanings when they vocalized, despite not having any kind of visual contact with their respective partners. These examples illustrate that pantomime *can* accompany speech since there is no physiological or cognitive constraint prohibiting a hearing speaker from gesturing and vocalizing simultaneously. The simultaneous surfacing of iconic gesticulations with speech should not presuppose that these gestures are fundamentally different from pantomime, or that the two followed different evolutionary trajectories; at the most basic level, they both reveal the natural tendency of humans to represent things mimetically. Pantomime and co-speech gestures share something that probably overrides any alleged differences; they are both propelled by our ability to entertain concepts which are themselves instantiated in our brains as mental representations (Brown in prep. a). Early hominins could have surely employed whatever means they had at their disposal to communicate; presupposing of course that their cognitive capacity had reached the level required to undertake more human-like communication. It might be the case that pantomime was just more "handy". Thus, we should not discard the possibility that pantomime can play an important role in establishing a communication system that has both arbitrary and iconic elements.

[5] It was pointed out to me that this is also quite similar to what children do while playing (M. Tamariz, personal communication, July 04 2011).

3. The gestural hypothesis of language origins

In examining gestures and its relation to language it is necessary to briefly review some of the most prominent gestural hypotheses of language origins. Corballis (2003) stresses that primate vocalizations cannot be viewed as precursors to human language due to their holistic and involuntary nature; in terms of evolutionary steps needed to be taken towards human language, gestures appear as a better candidate to undertake this shift. Among the fundamental features of gestural theories of language evolution is the idea that speech can be viewed, as several authors claimed (Neisser 1976; Studdert-Kennedy 1987; Browman & Goldstein 1991), as composed of articulatory gestures taking place within the mouth rather than as combinations abstract linguistic units – phonemes. Some authors seem to be arguing for the fact that a fully-fledged language evolved initially in the gestural mode, and was subsequently "translated" into speech (Armstrong et al. 1994; Corballis 2003, 2010), while others believe that gestures paved the way for speech without necessarily having formed a language-like (i.e. in terms of grammar) communication system first (Arbib 2005; Tomasello 2008; Zlatev 2005, 2008). Evolutionary linguists are also at odds over whether this ancestral "language" included arbitrary and/or conventionalized forms, or not.

Armstrong et al. (1994, p. 351) argue in favour of a gestural theory of the origins of syntax. Due to the fact that the human hand took its modern shape relatively early in evolution – compared to the vocal tract which presumably evolved much more recently and exclusively within anatomically modern H.Sapiens – "visible gestural communication took the early lead with respect to flexibility and complexity of output". This is not to say that vocalizations were not important from the very beginning of language evolution, but rather that analysing the physical structure of iconic manual gestures allowed early humans to make the first step towards a syntactic language (ibid.). Thus, Armstrong and colleagues (1994) gestural language is composed by arbitrary and conventionalized gestural forms organized around a human-like grammar. As the authors put it, "to see the rudiments of syntax it was necessary only to look perceptively at the parts of gesture such as grasping a finger or miming throwing; "getting to syntax was a matter of analysis, not synthesis" (p. 353). Nevertheless, Armstrong and colleagues' (1994) answer to the "translation problem",

i.e. why did speech become the dominant mode of linguistic expression, remains quite problematic. Claiming that any language (spoken or sign) is fundamentally based on gesture is not enough to explain the reasons why hominins needed to shift away from a fully functional communication system to another one.

Corballis (2003, 2010) agrees with Armstrong et al. (1994) that a full gestural language could have preceded the transition to speech. However, like the aforementioned authors, he fails to provide a persuasive account of a smooth transition 'from hand to mouth'. Corballis (2010) is not particularly interested in syntax, but rather bases his claims on the discovery of mirror neurons in monkeys (area F5), and their putative homologues in the human brain (Broca's area; BA44, BA45). Supporting a scenario which encompasses moving from pantomime to symbolic and conventionalized gestures (both brachio-manual and oro-facial), and finally to speech, Corballis (2010) suggests that facial and mouth gestures played a transitional role in language evolution. Hence, at some point, mouth gestures were dubbed with a vocal element which augmented their accessibility and distinctiveness (e.g. voiced vs. voiceless sounds) (ibid.).

Arbib (2005), Tomasello (2008) and Zlatev (2005, 2008) are three more supporters of the gestural hypothesis of language origins. Their ideas are essentially akin, with Zlatev's (2005, 2008) scenario diverging in a slight, but nevertheless highly important way, from those proposed by Tomasello (2008) and Arbib (2005). In contrast to the aforementioned theories, these three authors' evolutionary scenarios have a rather multimodal character. Arbib (2005), expanding on his "Mirror System Hypothesis"[6](see Arbib 2003), claims that the two intermediate stages in his model – 'protosign' and 'protospeech' – evolved in an expanding spiral, "so that the full development of protospeech would be impossible without the protosign scaffolding" (p. 158). He believes that the capacity for pantomime (part of 'protosign' stage), recruited by hominins due to its inherent communicative nature, involved a "genuine neurological change" in the hominin mirror neuron system (p. 158). Pantomime was the fundamental stone in the establishment of 'protosigns' (ibid). Using pantomime, early hominins were able to form a vocabulary of arbitrary and conventionalized gestures

[6] "The parity requirement for language in humans – that what counts for the speaker must count approximately the same for the hearer – is met because Broca's area evolved atop the mirror system for grasping with its capacity to generate and recognize a set of actions" (Arbib 2005, p. 149)

used to disambiguate the meaning of the original pantomime (this is where Zlatev (2005, 2008) disagrees, as I will explain below). It was the ability of employing arbitrary symbolic gestures to communicate new meanings that acted as a stepping stone for 'protospeech' (ibid.). Consequently, this 'expanding spiral' resulted in a language-ready brain, regardless of the sensory modality providing its input. Arbib's (2005) proposal is still open to the question of "why did speech predominate if protosign was so successful"? He tries to answer this by claiming that if protolanguage resulted from the interweaving of protosign and protospeech, then speech did not actually replace a full gestural language but a system built upon a "vocal-manual-facial complex" (p. 165). Thus, the apparent human "bias" for speech can be explained on historical terms rather than biological ones (ibid.). These arguments can explain, at least to some extent, why sign languages are acquired as easily as spoken ones, as well as why gestures accompany speech in everyday communication; what McNeill (1992, 2005) calls 'gesticulation'.

In a similar vein, Tomasello (2008) argues that human communication originated from the same "psychological infrastructure that is present already in species-unique forms of gesturing such as pointing and pantomiming" (p.218). He believes that the first arbitrary conventions emerged out of cooperative iconic gestures which through historical time become increasingly arbitrary and standardized. Tomasello (2008), like Sperber (2000), supports that language need not, or better, could not have preceded our ability to understand the other people's intentions. It is rather the latter, what Sperber (2000) calls our "metarepresentational capability" – embedding representations within representations, and more specifically, intentions within intentions – which should have acted as a springboard for the evolution of language; the two of them subsequently co-evolving in an evolutionary spiral. This species-unique ability allowed iconic gestures to function as the first communicative acts, galvanizing the shift toward more language-like communication[7]. The combination of iconic gestures with other iconic or pointing gestures could have added additional expressive power to this simple communication system (ibid.). The switch to the vocal modality happened basically through historical time, and was the result of using

[7] The same argument was used in the final paper I submitted in the 'Evolution of communication' course (Semester 2)

vocalizations along with already symbolic and conventionalized gestures; at some point gestures and vocalizations expressed the same meaning, and due to the advantages the latter offers in communication (freeing the hands, communicating at longer distances, etc), gestures become redundant.

McNeill and colleagues (2008) challenge the explanatory power of gestural theories, the likes of Corballis (2003, 2010), Arbib (2005), and Tomasello (2008), in dealing with the gesture-speech dialectic inherent in human language. As mentioned above, speech-synchronized gestures are an indispensable part of human communication and not only supplementary to conventional linguistic forms (McNeill 2005); imagistic gestures carry information not easily communicated by language, whereas language carries most of the referential meaning. McNeill and colleagues (2008) underline the discussed theories' inability to explain how an "integrated system of synchronized gestures and spoken forms" actually evolved (p. 1). They believe that the evolution of pantomime does not reflect the end-state of language origins (ibid.). According to their account, even if hominins initially used pantomime to communicate, language, as we find it in modern humans, must have followed a different evolutionary path at a subsequent stage. This is justified by the aforementioned different temporal relationship gesticulation and pantomime share with speech (synchronous vs. alternating) (ibid. p.10). McNeill (2005) may be correct in highlighting a difference between these two different forms of gestures, but we should not hastily discard the possibility of a link between them in the course of language evolution.

Zlatev's (2005, 2008) theory of language evolution is based on the idea of "Bodily Mimesis", which the author considers the "missing link" in human cognitive evolution. According to it, a particular bodily act of cognition or communication is an act of bodily mimesis if and only if:

(a) It involves a cross-modal mapping between exteroception (i.e. perception of the environment, normally dominated by vision) and proprioception (perception of one's own body, normally through kinesthetic sense).

(b) It is under conscious control and corresponds – iconically or indexically – to some action, object or event, while at the same time being differentiated from it by the subject.

(c) The subject intends the act to stand for some action, object or event for an addressee (and for the addressee to recognize this intention).

(d) Without the act being conventional-normative.

(e) Without the act dividing (semi)compositionally into meaningful sub-acts that systematically relate to each other and other similar acts. (Zlatev 2008, p. 138)

Similarly to Arbib (2005), Zlatev (2005, 2008) constructs a "layered model" in which every new capacity builds upon a previous one without replacing it and with all layers operating in tandem.

The feature that differentiates this hypothesis from Arbib's (2005) is that Zlatev's (2005, 2008) 'triadic mimesis' stage – the alleged "missing link" between human and nonhuman primate mimetic skills – is not equivalent to the 'protosign' stage in Arbib's (2005) theory. This is because pantomime, according to Zlatev (2005, 2008), and contra Tomasello (2008), was not symbolic or conventionalized in any way. Instead, it had to be "re-negotiated on a case-by-case basis" (Zlatev 2008, p. 146). Its communicative power derived from the fact that a pointing gesture could have been combined with an iconic one, forming simple predications (ibid.). Zlatev (2008) finds this to be comparable to what we find in 'homesigns'. The term 'homesign' refers to the gestural communication systems of deaf children who lack normal linguistic input[8]. Deafness prevents these children from being exposed to the spoken language of their hearing relatives, plus they are not exposed to a sign language during their first years of their lives (Goldin-Meadow& Mylander 1990). Nevertheless, they are still able to find a way to communicate, mostly with their families, by employing a combination of iconic ('pantomime') and pointing gestures (ibid.). Being fundamentally a process of language creation *de novo*, homesigns appear to fit quite nicely the situation just described for our hominin ancestors; albeit early hominins might have lacked the brainpower and learning biases human infants are genetically equipped with, as well as a highly enculturated environment in which they could develop their communication skills (Botha 2007).

Zlatev (2008, p. 146) claims that in a subsequent stage linked with protolanguage speech might have been initially employed to ""disambiguate"...stabilize and *conventionalize*

[8] The same argument based on homesigns was used in the final paper I submitted in the 'Biolinguistics' course (Semester 2)

(Zlatev's italics) meanings". He believes that this rationale avoids the pitfall of a hypothesized mirror system which can encompass "the understanding and sharing of actions but not of signs" (ibid., p. 148 citing Hurford 2004). Hence, the proposed scenario goes as follows:

(a) Moving from predicational non-arbitrary and not fully conventional triadic mimesis, i.e. pointing-to-referent + miming-an-action, **TO**

(b) Predicational, but multimodal and conventional protolanguage, i.e. object-sign + action-sign

The purpose of this paper is not to decide which of these theories is right or wrong. What appears fairly persuading though, both in Zlatev's (2005, 2008) and Tomasello's (2008) line of argumentation, is that arbitrary linguistic communication in the vocal-auditory channel could not have evolved from scratch. The attempt will be to show that gestures (pantomime) *can* serve as a bootstrapping device between meanings and arbitrary sounds – even in the absence of feedback – if the latter are repeatedly associated with iconic gestural signs in the context of reciprocal pair-wise interactions.

4. Garrod et al. (2007) experiment

The experiment presented in this paper was inspired by the way Garrod and colleagues (2007) examined the evolution of graphical communication systems. In their experimental setup, participants were placed into pairs having to play a game similar to the parlour game 'Pictionary'. One of the participants (the matcher) had to identify a concept drawn by her partner (the director) from a set of 16 alternatives. The experiment went through 6 consecutive rounds (blocks) each consisting of 12 of the 16 available concepts. The concepts were drawn from a fixed list of easily confusable items (e.g. museum, theatre; Robert de Niro, Clint Eastwood; television, microwave, etc). This experiment involved several conditions, drawing comparisons on many different levels. Initially, Garrod et al. (2007) compared how graphical communication systems evolve under three separate conditions; (a) one in which a Single Director with no Feedback (SD-F) drew items for an imaginary audience; (b) one in which a Single Director with Feedback (SD+F) draw items to be identified by the matcher (her partner); and (c) one which involved a Double Director with Feedback (DD+F), i.e. the participants alternating roles between rounds as matchers and directors (from these three conditions, the last one (DD+F) is the most relevant to my experiment and will be emphasized the most).

The results revealed that identification accuracy – the number of correct guesses a participant was able to make based on the director's drawings – increased across rounds for all three conditions, the rates, however, being lower in the SD-F condition than in the SD+F and DD+F conditions. This difference was statistically significant when comparing the SD-F with the SD+F conditions, and the SD-F with the DD+F condition, but not when comparing the SD+F with the DD+F condition. Examining the graphical complexity of the participants' drawings in all three conditions, the experimenters claim that in the SD-F condition, involving merely repetition and no interaction, drawings tended to become increasingly more complex and retained their iconicity. Conversely, in the other two conditions, involving interaction, drawings, according to the authors, tended to become more symbolic

and less elaborate[9]. It is noteworthy that participants in the DD+F condition improved slightly at a faster rate than those in the SD+F condition, pointing to the possibility that reciprocity in the interaction might enhance processes of simplification and refinement through grounding.

The authors conclude that interaction – a combination of imitating the signs of one's partner, and of providing feedback – is critical for the shift from iconic graphical signs to more symbolic ones. However, it could be argued that the symbolic status of this kind of signs is, at least, debatable since there is no indication in Garrod et al.'s (2007) paper of how the participants themselves perceived and interpreted their signals; we should not exclude the possibility that the signs carry within them a great deal of residual iconicity and that they were still iconic for the people who created them (despite being less iconic to an "outsider") (Brown in prep. b). Brown (in prep. b) notes that defining a sign's status as symbolic, iconic or indexical is highly subjective since we must take into account the sign's context-of-use. In our effort to discern what is a sign we have to consider not only the relationship between its meaning and its form but also the relationship between these two and the sign's user; "...without reference to the user, we cannot know how form relates to meaning" (ibid., p. 2). In order to be certain if people are using a sign arbitrarily (i.e. as a symbol) or non-arbitrarily (i.e. as an icon or index) we have to know how they perceive its form in a given context (ibid.). This was the main reason for deciding to use a post-test questionnaire, assuming it would provide insight into how participants perceived the signs they created.

[9] The authors also examine how the degree of interaction affects the evolution of signs, as well as how people not directly involved in the interaction deal with recognising the signs created by participants in the SD+F and the DD+F conditions. Since these are not directly related to my experiment, and comparisons will not be able to be draws, they will not be discussed even further.

5. Experiment

5.1. Experimental Design

The aim of this experiment was to model a situation where a multimodal and semiotically mixed communication system evolves through social collaboration. The experiment adapted Garrod et al.'s (2007) experimental design which, as it was previously said, examines the evolution of iconic graphical signs into symbolic graphical signs through repeated usage or processes of grounding. The present design, however, examines how the same process takes place if we replace the graphical element with a gestural one (i.e. pantomime). More specifically, it examines how gestural and/or vocal communication systems evolve through repeated pair-wise interactions. Thus, it will be shown that in the presence of multiple channels of communication (gestural & vocal) a shared, arbitrary sign system can emerge.

The experimental setup was similar to the parlour game "Taboo: Body Language" (charades) (Garrod et al. 2007 used "Pictionary") in which people try to convey certain prescribed meanings to their teammates. As in the Garrod et al.'s (2007) experiment the concepts to be indentified were drawn from a fixed list that was publicized to all participants. The list comprised of 12 distinct meanings grouped intentionally in four triads of easily confusable items (see APPENDIX A, Pic. 1 & 2 for images of objects). The difficulty level of gesturing each item independently was estimated according to its 'manipulability' (scaling from 0 for easily gesturable items to 4 for hard-to-gesture items). The 'manipulability' ratings used in the experiment are taken from Brodeur et al. (2010), and basically describe how easily people believe they can gesture what an object is to another person (using a scale of 1 to 4, with lower numbers being harder to gesture and lower ones being easier). In my experiment, half of the items were chosen for having 'low manipulability', i.e. being harder to manipulate (manipulability $\leq$ 2.1), whereas the other half were chosen for having 'high manipulability', being easier to manipulate (manipulability $\geq$ 2.9). The selected items were rather simple and straightforward since the experiment's purpose was not to test the participants' abilities in playing charades (although variation in performance is still unavoidable); this was also the reason for including only

items from a single category (i.e. objects) instead of from the variety of categories (places, people, programmes, objects, abstract) Garrod and colleagues (2007) included in their study.

'Low Manipulability' items

Target item	Item category	Manipulability
Pepper	Food	1.2
Orange	Food	1.8
Staples	School supply	2.0
Speaker	Electronic	2.0
Lemon	Food	2.1
Alarm clock	Electronic	2.1

'High Manipulability' items

Target item	Item category	Manipulability
Water bottle	Sport	2.9
Wine bottle	Food	2.9
Tape recorder	Electronic	2.9
Hole punch	School Supply	2.9
Champagne bottle	Food	3.2
Stapler	School supply	3.5

Table 1: List of items divided to 'low' and 'high' manipulability items

Target item	Item category	Manipulability
Staples	School supply	2.0
Hole-punch	School supply	2.9
Stapler	School supply	3.5
Speaker	Electronic	2.0
Alarm clock	Electronic	2.1
Tape recorder	Electronic	2.9
Water bottle	Food	2.9
Wine bottle	Food	2.9
Champagne bottle	Food	3.2
Pepper	Food	1.2
Orange	Food	1.8
Lemon	Food	2.1

Table 2: List of items divided to 4 groups of easily confusable items

Participants were asked to communicate either through gesture[10] ('gesture-only' condition), or through a combination of gesture and sound ('gesture-and-sound' condition), through a series of 4 rounds in which all of the target meanings were to be communicated. A significant difference with the original experiment is that my experiment involves merely a Double Director with no Feedback (DD-F) condition. All sessions were recorded by a digital video-camera capturing both participants in either side of the obstruction to facilitate subsequent analysis.

5.2. Pilot Study

An informal pilot study was conducted to ensure that the methodology and materials used in the experiment were sufficient to elicit useful and analysable observations. Participants in the pilot study were two monolingual Greek-Cypriot natives (recruiting English natives as in the actual experiment was considered trivial since what was meant to be checked was merely how the experiment would progress across rounds and not to actually draw any conclusions about my experimental hypotheses). The exact same materials and procedure were used for conducting a full game of the 'gesture-and-sound' condition (5 rounds in total). This was done in order to estimate the approximate time the experiment takes, and with it the amount I needed to pay each participant. The pilot study lasted approximately 75 minutes. The pilot was run smoothly and with no specific problems leading to the decision that modifications in the experimental design were unnecessary. For the record, the two participants were quite successful in playing the game. This hinted to the possibility that at least some of the participants in the actual experiment would be able to attach some meaning to the sounds given to them through a grounding process founded on gestural communication.

5.3. Participants

Participants were 28 monolingual speakers of English (10 males – 18 females, age 19-37). 14 of them took part in the 'gesture-only' condition, and the remaining 14 in the 'gesture-and-sound' condition. In both conditions participants were divided in 7 pairs. Participants were recruited through the University of Edinburgh's webmail service and

[10] The term 'gesture' in the experiment is used to include movements with any part of the human body (hands, arms, legs, orofacial expressions) and should be read as this unless told otherwise.

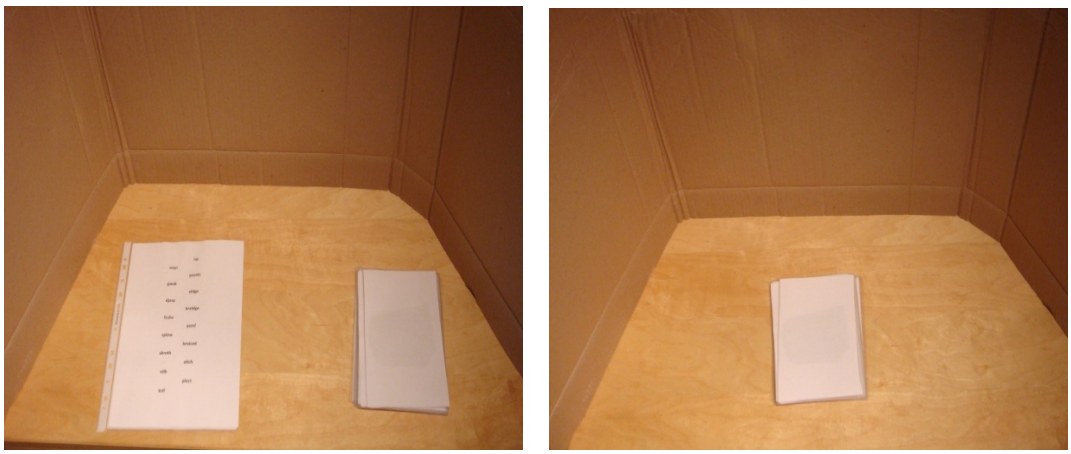
through the 'Student and Graduate Employment' (SAGE) database. Each participant was paid £9.

5.4. The 'Gesture-only' Condition

In the 'gesture-only' condition participants were asked to communicate merely through gestures. Upon arrival, they were given printed instructions describing the game as well as the rules they would have to follow (see APPENDIX A for game's instructions sheet). Nevertheless, they were not given any information about the purpose of the experiment, or about my experimental hypotheses. 14 monolingual English natives were randomly grouped into 7 individual pairs and were asked to play 4 consecutive rounds of the game with their partners. Each game consisted of 24 trials in which each pair played each item twice but in a random order (ensuring in this way that participants would be unable to use the order of play as a strategy for guessing the intended meaning). Thus, all pairs played each item 8 times in total. If we take into consideration that this experiment involves a kind of double rounds (2 times per round x 4 rounds = 8 rounds) then we end up with two more rounds than in Garrod and colleagues' experiment (6). This interaction time was considered sufficient to allow the elicitation of useful observations.

Before starting off, two different sheets including all the items in coloured printed form, but in a different random order (see APPENDIX A, Pictures 1 & 2), were given to the participants (one for each) allowing them to familiarize themselves with the items. If they had any uncertainties about the items or the game's rules they could ask the experimenter for clarifications. The interacting pair was placed in the same room; however, the two participants were separated by an obstruction created by two large carton boxes. The required materials for each participant were laid out on each side of this obstruction. The participant who performed the gesturing sequences (the 'director') had in front of her a stack of 24 small, laminated cards with the pictures of the target items printed on them (each item x 2 cards) (see Picture 1). The participant who had the guessing role (the 'matcher') had in front of her an array of the 12 cards depicting all the items to be communicated (see Picture 2).

A round ended once all of the target items were communicated and no card was left in the director's stack. Before the beginning of each round, the cards were shuffled by the experimenter, and they were placed upon the table face down. The director was then asked to draw the card that has ended on top of the stack. Thus, in each round, items were communicated in a different random order. Directors were asked to turn the selected card towards the video-camera situated in the room, ensuring that all data would have been available to analysis. After seeing the target item, they had to put it aside and reveal themselves to their partner by moving away from the obstruction. They could then start gesturing out the item presented on the card, using any body part they wanted but without using any kind of vocalizations. Their partners (the 'matchers') had to guess the concept that the 'director' intended to communicate by choosing one of all of the possible items included on the cards in front of them. Once the matchers thought they understood the intended meaning, they could tell their partner to 'STOP' gesturing, pick up the card including their guess and show it to the camera (see Picture 3). Matchers could choose whichever card they thought matched their partner's sign and they could choose any card as many times as they wanted (e.g. someone might choose to a single card as her guess for all trials). The target items and the matchers' guesses were simultaneously noted down by the experimenter on a separate answer sheet ensuring all answers would have been available for analysis (see APPENDIX A, Tables 2 & 3 for Answer Sheets in printed form).



Picture 1: Director's materials in the Gesture-only (Left) and in the Gesture-and-sound Condition (Right)

Picture 2: Matcher's materials (Both Conditions)

Picture 3: Starting from the upper-most image, the director (on the right) picks up the card form the stack of cards and shows it to the video-camera. At the same time the matcher (on the left) is still situated behind the obstruction so as not to see the target item. He subsequently performs a gesture to communicate the target meaning. Once the matcher thinks he has understood the intended meaning he tells the director to stop. The director moves in a position so as not to see his partner's guess, and the matcher shows his answer to the video-camera

Similarly to the Garrod et al. (2007), this experiment involved a 'Double Director' (DD) condition. Nevertheless, no feedback was available, apart from the fact that the participant gesturing was in a position of assuming that her actions were at a satisfactory

level for allowing her partner to make a guess (i.e. by stopping the director, the matcher indirectly communicates she has probably found the target meaning). More specifically, feedback exchange was not allowed neither in the form of queries followed by corrections and modifications in the original sign – what was allowed in Garrod et al.'s (2007) experiment – nor in the form of revealing their intended meaning or guesses to each other (based on the role they had in the interaction, i.e. director / matcher). Hence, the pair was not informed about the success of its interactions. Stopping one's partner once a matcher thought to have found the correct target meaning and the reciprocity of the interaction were considered enough to compensate for the absence of feedback.

Another important reason for excluding the possibility of feedback was to highlight the communicative power of gesture and its great potential in acting as a bootstrapping device between meaningful concepts and meaningless sounds. In the absence of feedback, trying to communicate merely through arbitrary sounds would have been virtually impossible. Vocalizations are iconically weaker than gestures since they cannot resemble their referents as efficiently as gestures can (Brown in prep. a). Thus, they are not as good as gestures to establish communication when a shared code is absent. (A 'sound-only' condition was excluded from this experiment due to time limitations but also due to the logical assumption that participants would have been unable to communicate by means of arbitrary vocalizations unless they could exchange feedback. Nevertheless, this is an assumption that can and needs to be verified in future experiments).

Within pairs, partners switched roles, acting as 'directors' in half of the trials, and as 'matchers' in the other half. Hence, each participant gestured a total of 48 items (2 Rounds x 24 items) and guessed another 48. More specifically, after the end of Round 1 the cards in the director's side were reshuffled (the same applying for the array of cards in the matcher's side) and participants swapped roles by moving to the alternative side of the obstruction (see Table 3). At the end of Round 4, participants were asked to fill out a questionnaire describing their game experience (strategies they followed, reasons underlying their choices, etc). Participants in this condition had to answer only 2 out of the five questions included in the questionnaire since only those applied to the version of the game they played, whereas participants in the second condition had to answer all 5 of them (see APPENDIX A for

questionnaire form). The main reason for including a questionnaire was mainly to have a better grasp of the iconic status of gestures and if they were used as a scaffold for giving meaning to sounds.

	Round 1		Round 2		Round 3		Round 4	
TRIAL	Director	Matcher	Director	Matcher	Director	Matcher	Director	Matcher
1	Par.A	Par.B	Par.B	Par.A	Par.A	Par.B	Par.B	Par.A
2	Par.A	Par.B	Par.B	Par.A	Par.A	Par.B	Par.B	Par.A
3	Par.A	Par.B	Par.B	Par.A	Par.A	Par.B	Par.B	Par.A
. . . 24	Par.A	Par.B	Par.B	Par.A	Par.A	Par.B	Par.B	Par.A

Table 3: Game procedure – Participants playing 4 rounds in pairs, alternating roles (directors / matchers) across rounds

5.5. The 'Gesture-and-sound' Condition

In the 'gesture-and-sound' condition, participants of a separate group (7 more pairs) were asked to communicate the target meanings by using a combination of gesture and sound (novel English non-words), the procedure being almost identical to the 'gesture-only' condition. Employing both gesture and sound was mandatory; instead of giving participants the choice of using the novel sounds if they wanted or not. However, this factor should not be seen as a limitation of the experiment. The experiment intended to examine what would happen if non-arbitrary sounds accompanied non-arbitrary gestural communication. This is based on the premise that human communication is fundamentally multimodal (McNeill 2005, Brown in prep. a). Consequently, a gesture – sound complementariness is something to be expected; the way the experiment is set merely manipulates the vocalization component in order to maximize arbitrary interpretations (E. Brown, personal communication, June 30 2011).

The sounds were drawn from a set of distinct non-existing sound strings following English phonotactics. The word list was deliberately bigger than the object list in order to prevent participants from successful guessing by means of some kind of elimination processes. The word list was generated through the ARC Non-word Database and the novel words were chosen on the basis of having a low level of phonological neighbourhood

(phonological neighbours ≤ 2) (see Table 4). All novel words were composed of 3 – 7 letters and 3 – 6 phonemes in order to be easy to remember and pronounce.

Before starting the game, participants were given the opportunity to familiarize themselves with the pronunciation of the selected novel words. This was done by giving them a list including all words in a written form and by exposing them to a two-minute sound clip including an audio recording of all sounds pronounced by a native English speaker (the words in the list and the sound in the audio file had the same order, see APPENDIX A, Table 1 for the list's printed version). All sounds were produced twice and were divided by brief intervals allowing the participants to attempt pronouncing the sounds themselves. It was made clear to the participants that the novel words had no association with any of the target concepts, or with any pre-existing English word.

Apart from the stack of cards, the director had now access to a sheet including the words in a written form (a different one from the one they were given in the practice round, see APPENDIX A, Pictures 3 & 4). In order to communicate a target meaning, the director was allowed to use any kind of gesture, plus one of the words included on the list. No explicit instructions were given for when to produce the sound (participants were free to use the sound before, after, or during their gestures). Similarly to the 'gesture-only' condition, the matcher had to call her partner to stop as soon as she thought she was in a position to guess correctly, but not before the director had produced *both* the gesture and the sound. After the end of each round, a different sheet including again all the words but in a different order was placed in front of the new director (this was done to ensure that the pair would not use this list as a means of establishing a communication strategy based on the items' order).

Instead of playing 4 rounds of the game, participants in this condition played an additional 5th round in which they were asked to use merely the sounds, but no gestures, in order to communicate. In this round, the interacting pair played again all the 24 cards, with one of them acting as a director in the round's first half (12 trials) and as a matcher in the second half (12 trials). In other words, both participants had the opportunity to produce a total of 12 sounds in order to communicate 12 cards, and guess what the sounds meant in another 12 trials. It should be noted, that the participants were not informed about the

existence of this extra round until they had played the last card in the 4th round. Consequently, they were not aware that my goal was to check whether they would attach meaning to the available sounds. This allowed us to check if the pair had established a communication code in which sound and meaning have an arbitrary relationship.

Non-word	Pronunciation	*N* of Phonological Neighbours
krif	[krɜːf]	0
gwuk	[gwʌk]	0
yemf	[jemf]	0
vilb	[vɪlb]	0
yeents	[jiːnts]	0
altch	[æltʃ]	1
splow	[splaʊ]	1
skroth	[skrɒθ]	0

Non-word	Pronunciation	*N* of Phonological Neighbours
snen	[snen]	0
fruhv	[frʌv]	0
kreldge	[kreldʒ]	0
djorp	[dʒɒrp]	1
plect	[plekt]	2
oldge	[ɒldʒ]	0
isp	[ɪsp]	1
broiced	[brɔɪst]	2

Table 4: List of novel English words selected for low phonological neighbourhood

5.6. Experimental hypotheses

There were certain expectations for the development of signs as the game progresses across blocks. It was primarily expected that when people are found in an isolated pair condition leading to local interactions they will create local communication systems (sign or spoken) which will be shared within the interacting dyads but not across dyads which do not come into contact. Similarly to Garrod at al.'s (2007) experiment, it was expected that in both conditions: (a) partners would become increasingly better at identifying each other's signals; (b) their respective communicative signals would start to converge. Nevertheless, there was no expectation that the gestural signs would lose their iconicity and become more symbolic; this to be rather difficult to occur, especially under the conditions that apply in our experiment. A more symbolic gestural communication system might emerge in an experimental design involving a more dynamic interaction between the participants, something it will be briefly discussed in the paper's last section (e.g. creating communities of participants instead of interacting dyads). Gestures might lose some of their initial iconic

features due to their simplification, without necessarily claiming a symbolic status, i.e. they might still be iconic for the people who created them, and thus hold a non-arbitrary relation to their referents.

In the 'gesture-only' condition: (a) Gestures will become increasingly shorter; (b) When the participants alternate roles (director/ matcher) their gestures will progressively converge with those of their partners. With respect to the 'gesture-and-sound' condition it was expected that: (a) Participants will initially align upon a locally shared multi-modal and semiotically mixed communication system using both gestures (pantomime) and the available novel words to communicate (gestures will serve as a bootstrapping device between target meanings and arbitrary sounds); (b) Participants will eventually be able to use arbitrary sound-meaning pairings to communicate the target meanings and the new words will start to be conventionalized; (c) Gestures in this condition might become shorter if participants start to rely more on sound for communication than on gesture.

6. Measures for examining task performance

Gesture length/ complexity (Time)

Time was used as a measure to define gesture length / complexity required to communicate a particular gesture. In this respect, longer gestures are considered more complex, whereas shorter ones are considered simpler. The video-recordings of the experiment were used to define gesture length. The moment the director starts gesturing is the moment which marks the beginning of a gesture, while a gesture is considered to have ended once the matcher has called the director to stop gesturing[11]. It should become clear, that the exact method of measurement applies in the gesture-and-sound condition, i.e. if the director produced first the sound s/he selected and then started gesturing, the time needed for and following the production of the sound is not taken as part of the gesture (the same holds if the sound was produced after the end of a gesture). If the sound was produced simultaneously with the gesture then it was unavoidably counted as part of it.

Identification accuracy (Correctness)

For the two tested conditions (gesture-only, gesture-and-sound), identification accuracy was based on the performance of the interacting partners themselves. A correct answer was considered the one in which the director's and the matcher's cards matched in the context of the same trial. At the end of each round (24 trials) a running score was calculated for the pair (this was subsequently double-checked by examining the video-recordings). In the gesture-and-sound condition a separate score was taken for estimating the success of each pair in the additional 5th ('sound-only') Round.

Convergence on a shared code (Gesture Similarity)

Apart from the two abovementioned quantitative measures, 'convergence' (gesture similarity within pairs) was used as a qualitative measure to check whether participants align upon a locally shared communication system. This was achieved by examining the

[11] Gestures which lasted less than 1" were counted as lasting 1". Thus, 1" is the minimum gesture length achievable in this experiment.

obtained video recordings as well as the answers participants provided in the post-test questionnaires. This measure is specifically linked to the employment of gestures based on their apparent features and not on the use of sounds; however, it will become apparent that the convergence on a common code in the gestural mode is what permits convergence in the vocal-auditory one. If the gestures within pairs were similar / identical in the final round of the game, convergence to common code was considered successful. The post-test questionnaires were used to check whether participants consciously tried to employ signs similar / identical to those of their partners.

7. Results

7.1. Gesture length / complexity (Time variable)

To test the statistical significance of gesture length / complexity in the two conditions, both by-subject (F_1) and by-item (F_2) analyses were conducted. In the by-subject analysis a 4 x 2 x 2 mixed design ANOVA was run with 'Round' (1 – 4) (4) and 'Manipulability' (low and high) (2) as within-subject factors and 'Condition' (gesture-only, gesture-and-sound) (2) as between. In the by-items analysis again a 4 x 2 x 2 mixed design ANOVA was run with 'Round' (4) and 'Condition' (2) as within-subjects factors and 'Manipulability' (2) as between. The analyses revealed a highly significant main effect of Round both in the by-subject, $F_1(1.47,36)=65.39$, $p < .001$ (Huynh-Feldt correction); and in the by-item analysis, $F_2(1.88,30)=93.37$, $p < .001$ (Huynh-Feldt correction), as well as a significant main effect of Condition both in the by-subject, $F_1(1,12)=9.70$, $p = .009$; and in the by-item analysis, $F_2(1,10)=16.75$, $p = .002$. More specifically, participants in the gesture-only condition had significantly shorter gestures ($M_1=38.27$) than the participants in the gesture-and-sound condition ($M_1=51.39$) as the by-subject analysis reveals. The analyses did not reveal a significant main effect of Manipulability neither in the by-subject nor in the by-item analysis. Neither the by-subject nor the by-items analysis revealed statistically significant interactions between (a) Round and Condition; (b) Manipulability and Condition; (c) Round and Manipulability; and (d) Round, Condition and Manipulability. Further analysis of the Round effect (Bonferroni post-hoc test) revealed significant differences in the by-subject analysis (see Table 5) between the times of Round 1 ($M_1=79.36$) and Round 2 ($M_1=43.14$), $p < .001$; between Round 1 and Round 3 ($M_1=11.29$), $p < .001$; between Round 1 and Round 4 ($M_1=11.04$), $p < .001$; between Round 2 and 3, $p < .001$; between Round 2 and 4, $p < .001$; and between Round 3 and 4, $p = .005$. Equally significant differences were revealed in the by-item analysis (see Table 6) since the times of Round 1 ($M_2=92.46$) were significantly shorter than those of Round 2 ($M_2=50.13$), $p <.001$; those of Round 3 ($M_2=36.71$), $p <.001$; and those of Round 4 ($M_2=2.11$). There was also a significant difference in times between Round 2 and 3, p

< .001; between Round 2 and 4, p < .001. The difference between Round 3 and 4 was marginally insignificant, p = .052[12].

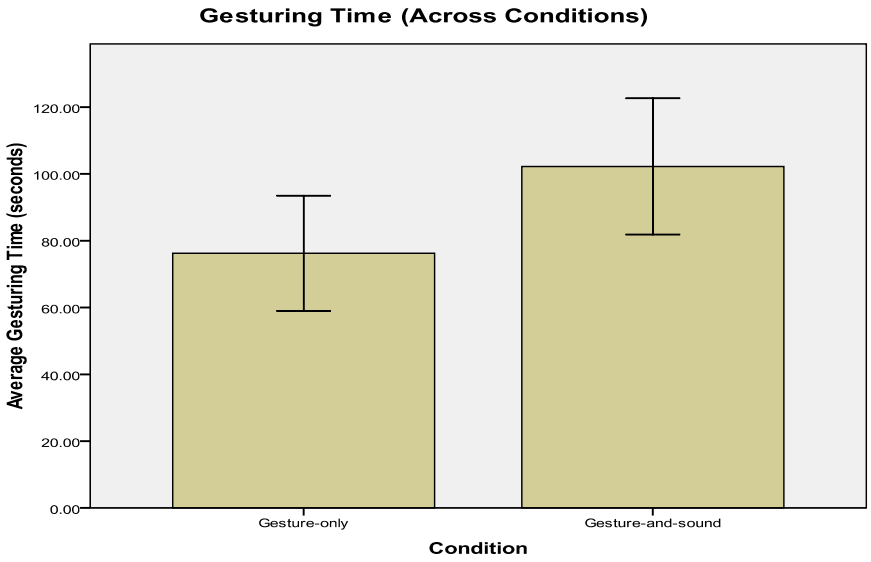


Figure 1: Chart showing the average time gesturing lasted in the two conditions

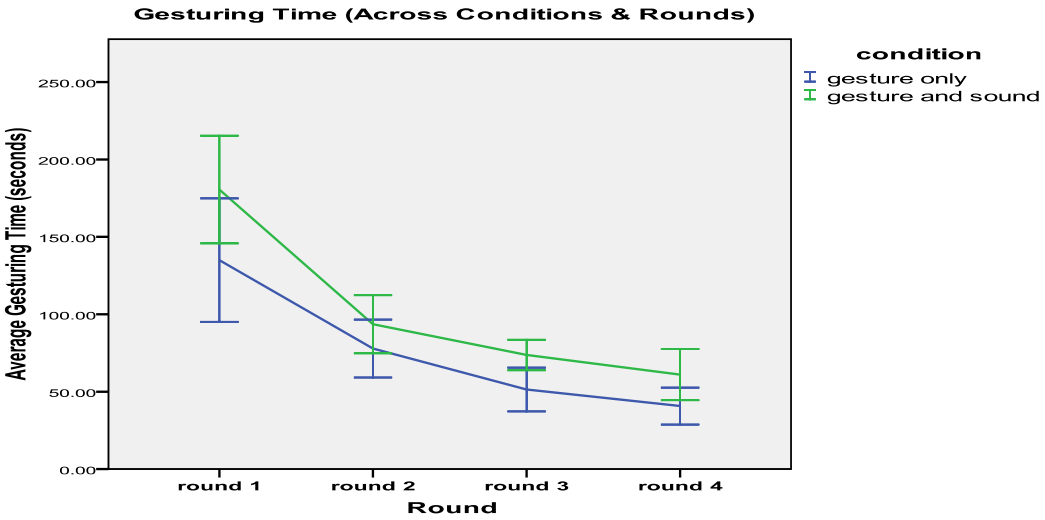


Figure 2: Chart showing the average gesturing time needed to complete a round across the two conditions

Round	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	79.357	5.420	67.548	91.166

[12] See APPENDIX D for additional tables with statistics.

2	43.143	2.735	37.184	49.101
3	31.464	1.804	27.533	35.395
4	25.393	2.095	20.828	29.958

Table 5: Mean Gesturing Time in Round 1 -4 (by-subject analysis)

Round	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	92.458	6.645	77.653	107.264
2	50.125	3.748	41.773	58.477
3	36.708	2.640	30.825	42.591
4	29.625	1.452	26.389	32.861

Table 6: Mean Gesturing Time in Round 1 -4 (by-item analysis)

7.2. Identification Accuracy

To test the statistical significance of the identification accuracy in the two conditions both by-subject (F_1) and by-item (F_2) analyses were conducted. In the by-subject analysis a 4 x 2 x 2 mixed design ANOVA was run with 'Round' (1 – 4) (4) and 'Manipulability' (low and high) (2) as within-subject factors and 'Condition' (gesture-only, gesture-and-sound) (2) as between. In the by-items analysis again a 4 x 2 x 2 mixed design ANOVA was run with 'Round' (4) and 'Condition' (2) as within-subjects factors and 'Manipulability' (2) as between. The analyses revealed a highly significant main effect of Round both in the by-subject, $F_1(3,36)=7.83$, $p = .000$; and in the by-item analysis, $F_2(3,30)=8.96$, $p < .001$, as well as a significant main effect of Condition both in the by-subject, $F_1(1,12)=6.29$, $p < .05$; and in the by-item analysis, $F_2(1,10)=15.48$, $p = .003$. More specifically, participants in the gesture-only condition performed significantly better in identification accuracy scores (M_1=11.52) than the participants in the gesture-and-sound condition (M_1=10.07) as the by-subject analysis reveals. The analyses did not reveal a significant main effect of Manipulability neither in the by-subject nor in the by-item analysis. Neither the by-subject nor the by-items analysis revealed statistically significant interactions between (a) Round and Condition; (b) Manipulability and Condition; (c) Round and Manipulability; and (d) Round, Condition and Manipulability (for all non-significant interactions – apart from Manipulability & Condition,

which is still non-significant – Fs < 1). Further analysis of the Round effect (Bonferroni post-hoc test) revealed a significant difference (p < .05) between the scores of Round 1 (M_1=10.04) and Round 3 (M_1=11.29) & 4 (M_1=11.04) in the by-subject analysis (see Table 7), as well a significant difference (p < .05) between the scores of Round 1 (M_2=11.71) and Round 3 (M_2=13.17), but not Round 4, in the by-item analysis (see Table 8)[13].

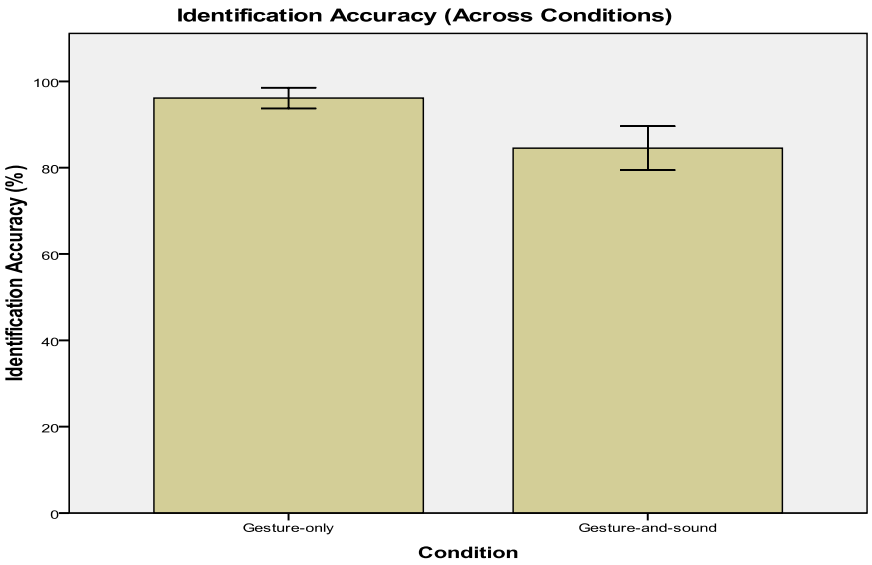


Figure 4: Chart showing the identification accuracy percentage across the two conditions

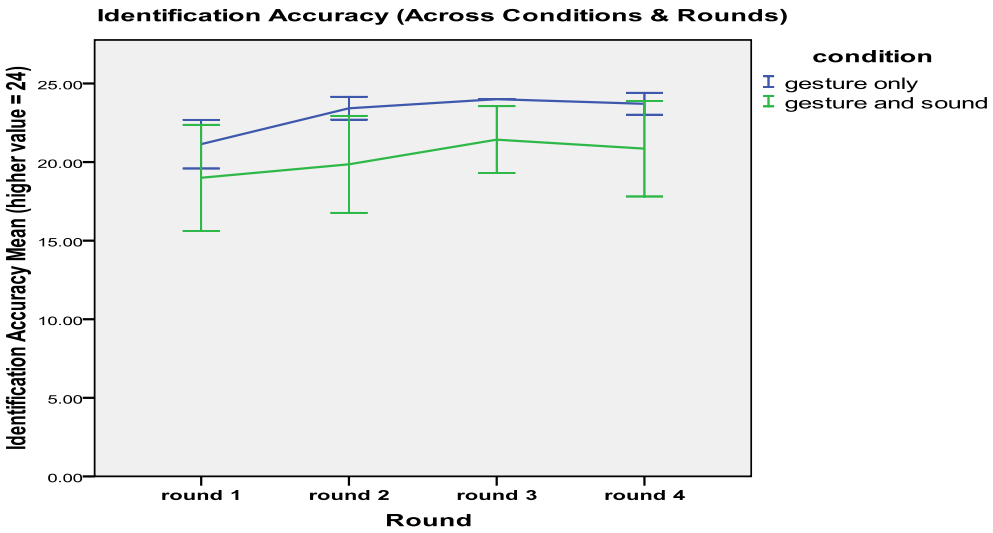


Figure 5: Chart showing the identification accuracy (number of correct answers out of 24) of the two conditions across the 4 rounds

[13] See APPENDIX C for additional tables with statistics.

Round	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	10.036	.380	9.209	10.863
2	10.821	.324	10.115	11.528
3	11.286	.251	10.739	11.832
4	11.036	.364	10.242	11.829

Table 7: Mean Identification Accuracy in Round 1 -4 (by-subject analysis)

Round	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	11.708	.510	10.572	12.845
2	12.625	.354	11.836	13.414
3	13.167	.286	12.529	13.804
4	12.875	.393	11.999	13.751

Table 8: Mean Identification Accuracy in Round 1 -4 (by-item analysis)

To check whether the identification accuracy in the final ('sound-only') round in the gesture-and-sound condition is significantly better than chance, a one-sample test was run using the scores of all 7 available pairs. The participants had to go through all the 24 cards available (each member of a pair being the director / matcher in 12 of them), and the participant who acted as a matcher had to choose which sound corresponded to which of the 12 cards in front of her (see back to section 5.5 for exact description). Similarly to the previous rounds, matchers could choose whichever card they thought matched their partner's sign and they could choose any card as many times as they wanted. Consequently, the test value for the one-sample t-test (i.e. the number of correct answers a participant might have been able to get simply by chance) was set to 2 (24 target object cards/12 possible matching cards = 2). The difference between the achieved identification and the level of chance was proven to be statistically significant, $t(6)=3.42$, $p < .05$. Even if we increase

the test value to 4 ('dummy' test value)[14], the result is still statistically significant, $t(6)=2.68$, $p < .05$ (the result ceased to be significant only when *test value*=5)[15]. This was done to counterbalance the possibility of the results being confounded by some kind of latent iconicity (onomatopoeia) or spelling similarity to existing words participants may have found in the novel words (as it will be discussed in section 8.3 it is rather unlikely that this factor have confounded the results).

Taking the scores of all pairs into consideration we end up with a mean score (*M*) of 11.29. It should be noted however that there is a great variability in the performance across pairs (*SD*=7.18). It is worth mentioning that some of the participants' answers in the last round were indeed the product of chance. This is proven by examining the video-recordings and the detailed answered sheets where it can be seen that a pair was able to communicate about a specific object with a sound they had never previously associated with this object (e.g. One participant using the word 'isp' to communicate 'lemon' even though 'isp' was never used when gesturing this particular item).

[14] I used a 'dummy' test value to examine the participants' success assuming they had to perform on an easier task (e.g. 24 target object cards / 6 possible matching cards = *test value* of 4). Thus, the exact same score they got in the game, in this case the number of correct guesses, is compared against a value that would make it easier to accept the Null-hypothesis that their success resulted from chance.

[15] See APPENDIX B for tables with statistics.

8. Discussion

8.1. Gesture length / complexity (Time variable)

By examining the graphs in Figure 1 and 2 (p. 34) comparing only Condition and Round (these were generated in order to have a more accurate graphing without considering the non-significant effect of Manipulability)[16] we can see that the gestures of participants in both conditions become increasingly shorter as they moved from Round 1 to Round 2, from Round 2 to Round 3, and from Round 3 to Round 4[17]. It is quite clear that the biggest difference is between Round 1 and 2. After that point the gestures continue to become shorter in both conditions but the point is not so pronounced. This to my opinion is mostly due to the fact that in Round 1 participants acted as directors had to invent their gestures from scratch. Apart from that the gesturing in Round 1 is much slower in terms of movement since directors probably wanted to ensure they would become understood – this applies for both conditions. In subsequent rounds, directors are much more fluent with their gestures; in the sense that they knew what they were doing, but not in the sense of gesturing in a "fast-motion" way.

Surely, this has to do with the fact that in Round 4 gestures in both conditions lasted on average below 3.4" (see Figure 3). Especially in the case of the gesture-only condition, many of the gestures could not have gotten any shorter in subsequent rounds (if those were played) since they lasted less than 1". It thus seems that as participants are able to interact for more rounds their gestures become significantly shorter. The key factor for allowing participants to simplify their gestures lies in the reciprocity of their respective interactions, i.e. that both act as directors and matchers. Hence, even in the absence of feedback, "simplification-by-grounding" as Garrod and colleagues describe it (p. 966) is still attainable. The obtained results suggest that the minimum feedback participants exchange by stopping their partners while gesturing (denoting something along the lines, "What you did was enough for me to find the right answer") is sufficient for this kind of simplification.

[16] Statistical analysis tables as well as additional graphs are included in the Appendix section.
[17] Figures 1 & 2 show how much time it took the participants to gesture out all of the target concepts and not how much time it took them to gesture a single concept.

The results are congruent with those of Garrod et al. (2007) who also found a significant effect for Round ('Block' in their experiment) revealing in their case that as the game progressed the graphical representation of participants become simpler and more refined.

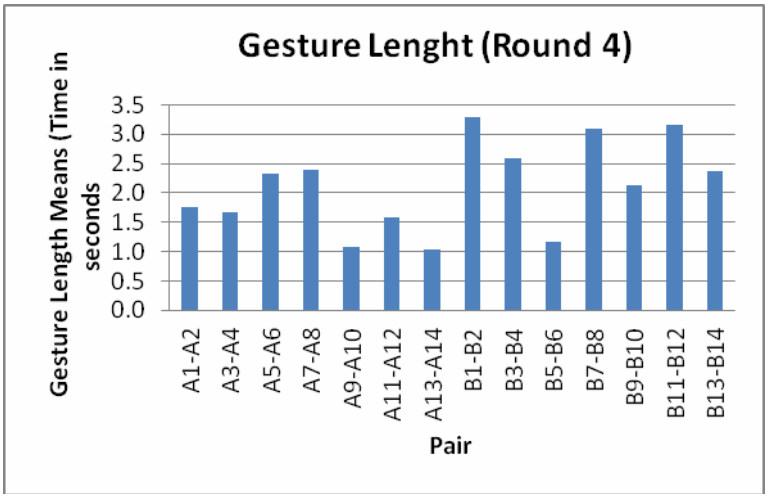


Figure 3: Chart showing the average gesture length (in seconds) of all pairs in Round 4 (A-A pairs come from the gesture-only condition, B-B pairs from the gesture-and-sound condition)

My hypothesis that gestures in the gesture-and-sound condition would become shorter than those in the gesture-only condition was not verified. This was based on the assumption that in the gesture-and-sound condition sounds would have gradually become the prominent medium of communication, and gestures would have become supplementary. On the contrary, gestures remained predominant throughout the experiment, while sounds were those assuming a supporting (or even trivial) role in communication. As the statistical analysis reveals, gestures in the gesture-only condition were significantly longer/ more complex than those in the gesture-and-sound. A plausible explanation is that in the latter, participants found it hard to remember which sound corresponded to which meaning. This might had an impact on the time matchers needed to identify the concept communicated by their partners. In Round 1, for example, matchers might have tried to find some kind of strategy that would have helped them remembering what sound their partner have used (if of course they were paying attention to sounds). In Round 2, directors might have expected their partners to produce the same gesture-sound combination they had themselves employed in the previous round. In facing a mismatch (i.e. same gesture-meaning pairing but different sound from what was used before), their reaction time might have been delayed since they would have to decide whether to keep this

sound or the sound they have used in Round 1 (this procedure could apply for all subsequent rounds or until the pair has come to an (indirect) agreement of what sound to use along which gesture).

8.2. Identification Accuracy

Examining the graphs in Figures 4 and 5 (p. 36) comparing only Condition and Round (similarly to the previous examination these graphs were generated excluding the non-significant effect of Manipulability) we can see that the performance of participants in both conditions improved as they moved from Round 1 to Round 2, and from Round 2 to Round 3, while it plateaus between Round 3 and 4 with a slight decrease in performance in the latter. It thus seems that when participants are able to interact for more rounds their accuracy in communication improves significantly. This happens because the kind of local interactions involved in the experiment produce accurate local sign systems that are shared within the interacting dyads. In agreement with Garrod et al. (2007, p.976), my findings seem to suggest that "reciprocal production and comprehension enhances the mutual intelligibility of the grounded...signs"; this applies both for the gesture-only and the gesture-and-sound condition, albeit less so for the latter. It should be noted that the graphs represent the overall mean performance of participants in different conditions; in the gesture-only condition all pairs achieved the maximum score in Rounds 3 and 4 (apart from one which scored 22/24 in Round 4), whereas in the gesture-and-sound condition existed similar but fewer cases (1 pair in Round 3 and 2 pairs in Round 4). This small drop (in both conditions) might be related to the fact that matchers in Round 4 tended to stop their partner's gesturing sooner than in Round 3 something that led to slight "misunderstandings" between easily confusable items which might had similar gestures. For example, by looking at Pictures 4 and 5 we can observe the apparent similarity between gesturing 'hole-punch' and 'stapler'. Both gestures involve creating an opening with the hands which then closes abruptly. In the case of 'hole-punch' the gesture is performed with both hands, while in the case of 'stapler' only the right hand is used (this led to misinterpreting the 'hole-punch' gesture as 'stapler').

Picture 4: Gesture for 'hole-punch'

Picture 5: Gesture for 'stapler'

What was unexpected was that the participants in the gesture-only condition would have performed better than the participants in the gesture-and-sound condition. The initial expectation was that both Conditions would have yielded similar results in terms of identification accuracy. A possible explanation for this discrepancy in performance is that the gesture-and-sound condition involved a more cognitively demanding task. Participants in this condition had to deal with some kind of extra cognitive/ memory load set on them by having to remember and use the novel words. This is not unrelated to the fact that gesture length was significantly higher in the gesture-and-sound than in the gesture-only condition, but unrelated to whether pairs used the sounds consistently or not. As far as directors are concerned, having to remember a sound might have hindered the ease of gesturing. On the other side, matchers might have delayed their decision to call their partners to stop gesturing by trying to spot some consistency in the sounds' usage, and by trying to save them in memory, knowing that in the next round they would play the game as directors[18]. This possibility becomes obvious by examining the experiment's video-recordings, with the

[18] Participants playing as matchers were instructed to wait their partner to produce both a sound and a gesture before they would tell them to 'STOP'. However, this could not have influence their performance since none of the participants produced the sound after gesturing (everyone did it either before or during).

flow of gestures of the gesture-only participants appearing much smoother and carefree. Conversely, the gestures of the gesture-and-sound participants are in the majority of cases introduced by short pauses marking their indecisiveness in choosing a sound.

8.3. Convergence on a shared code (Gesture Similarity)

Participants in both conditions were able to align upon a locally shared communication code. This is proven both by examining the video-recordings and the answers participants provided in the post-test questionnaires. The examination of the video-recordings reveals that this alignment entails the constant renegotiation of sign form between participants. Hence, Garrod and colleagues (2007) are probably correct in arguing that interaction promotes sign refinement (there are reasons to question, however, whether this leads to the transformation of icons into symbols – at least, in this experiment). The 'convergence' measure is firmly linked to the success of participants in correctly identifying their partners' gestures. In other words, convergence and identification accuracy go hand in hand in this experiment; in every instance in which pairs had similar / identical signs identification was successful (there were no instances in which participants within a pair had a different gesture for the same item and were still able to communicate, albeit this could definitely happen in another context). To illustrate how convergence and refinement occurs, it is useful to examine the pictures below and follow the establishment of a shared sign step-by-step.

In trying to communicate the concept of 'staples', participants A13 and A14 employed the gestures presented in Pictures 6 – 10. A13, starting the game as a director, initially tried to enact the action of getting staples into the stapler and then using it to staple some kind of imaginary paper (Picture 6). The series of gestures seen is Picture 6 was performed three times, and the whole gesture lasted 20" in total before participant A14 (the matcher) was able to provide a correct guess. As it was previously said, the object list contained easily confusable items, something that might have delayed the matcher from choosing, in this case, possibly between 'staples' and 'stapler'. The second time participant A13 encountered the same item, she performed the same series of gestures. A14 was able to guess correctly in 3", but this was not enough to preserve what her partner has introduced. Thence, in being a director herself, she decided to modify the 'staples' gesture (Picture 7).

Starting off with her right hand at a low position, she used it to iconically depict the image of ‘staple’, she then raised her right hand in the air and with her left hand pushed it down abruptly (gesturing the stapling action). This gesture got a delayed response (14") as well, since A13 might have expected to see what she introduced in Round 1. In the second trial of her Round 2, A14 used the same gesture, but using this time a refined version herself. Instead of starting off from below, her right hand was situated in the air from the beginning. The same stapling movement was enacted with the left arm, but this time she reminded to her partner that what she meant was the ‘staples’ by using again her right hand (Picture 8 – snapshot 3). As it can be seen in Picture 9, in Round 3, participant A13 decided to go along her partner and performed the same type of iconic gesture (in both trials of Round 3 this gesture lasted 2" and was identified correctly). In the final 4th Round (Picture 10), the use of the right hand was enough to lead to a rapid (1" in both trials) and correct answer. This example illustrates the reciprocity and dynamism of pair-wise interactions as well as the importance of iconicity in establishing common ground at the initial stages of communication; iconicity is what allows pairs to constantly renegotiate the meaning of a gesture (even in the absence of feedback), before reaching to a sign that will be used consistently to represent a particular concept.

Picture 6: Gestures for 'staples' Round 1 – Trials 1 & 2

Picture 7: Gesture for 'staples' Round 2 – Trial 1

Picture 8: Gesture for 'staples' Round 2 – Trial 2

Picture 9: Gesture for 'staples' Round 3 – Trial 1 & 2

Picture 10: Gesture for 'staples' Round 4 – Trial 1 & 2

Back to the question whether icons start to be transformed into symbols, my data seems to suggest that even though gestures get shorter and refined, they still maintain a non-arbitrary relationship with their respective referents. It might indeed be the case that the information starts to shift from the signs structure to the user's knowledge of the sign's previous use (Garrod et al. 2007, p. 964) since the participants' comments seem to suggest that what they gradually paid attention to was imitating their partners' gestures. Nonetheless, this does not mean that the signs are not still iconic. This argument is supported by observing several gestures for the same concepts across different pairs.

It was indeed surprising and contra the initial expectations, how similar (in some cases identical) in many cases the gestures of different pairs were (of course this is not unrelated to the fact that all of them shared the same cultural background). It could be alleged that the iconicity inherent in gestures somehow allows participants who have never interacted before to alignment upon a common (or at least highly similar) communication code. In Pictures 11-14, we can observe the apparent similarity in the way different pairs gestured 4 particular concepts; wine bottle, alarm clock, champagne bottle, and water bottle. For 'wine bottle', participants from different pairs (see Picture 11), used a gesture communicating the action of opening a wine bottle; one hand holding the wine bottle and the other hand used as a cork-screw (Note that all snapshots represent the final and abbreviated version of the gesture as it was used in the final round). For 'alarm clock', a gesture communicating the action of sleeping, interrupted by the sound of the alarm clock, was used (Picture 12). To communicate 'champagne bottle', the participants depicted in Picture 13, gestured how champagne bottles are usually opened by shaking the bottle and then pushing the cork which pops due to increased pressure. In the case of 'water bottle', the participants depicted in Picture 14, used some kind of running gesture to communicate the fact that after running, you can drink some water to avoid dehydration. Another gesture which was almost identical across the vast majority of pairs was the one used to communicate 'lemon'. This was a facial gesture expressing the sourness in tasting a lemon (due to confidentiality constraints pictures for this gesture were not included in the dissertation since it would necessitate revealing the participants' faces).

In all the aforementioned cases, the gestures have a non-arbitrary relationship to their referents, although it could be argued that the gesture for water bottle might be a little less transparent for an outsider than those for wine bottle, champagne and alarm clock. Despite being only abbreviated forms of longer gestures, we cannot claim that they have started to be transformed into symbols. Gestures can thus remain iconic and at the same time be shared between people who have never interacted before (i.e. being highly transparent). It may also be that pantomime is more conventionalised within a culture than we think (M. Tamariz, personal communication, July 04 2011). The prediction of this hypothesis would be that a cross-cultural comparison of the gestures for e.g. champagne bottle and water bottle would be more variable (ibid).

It is possible that the participants' responses in seeing these signs have become more automatic since repeated interaction has allowed them to save the signs in their memory. The discrepancy between the original experiment and mine might also partly result from the difference in the communication medium involved in the two experiments. In the original experiment, the fact that graphical representations can remain on paper might facilitate their transformation into symbols. Contrariwise, with gestures this is not possible. A gesture disappears as soon as the gesturer performs it leaving no tangible and observable trail behind it – it totally depends on the receiver's memory to remember the sign's configuration. It would also be interesting to see what would have happened if after the end of their respective games, participants were mixed up and reassorted into different pairs, being asked to use only the gestures they had used in the final round of their previous interaction. We could hypothesize that the non-arbitrary status of their gestures would have allowed them to communicate for the majority of items included in the object list (this is a hypothesis that can be tested in future experiments). Of course this kind of experiment would provide insight into a sign's 'transparency in form', but it would not tell us if the gestures were symbolic or iconic for the users (E. Brown, personal communication, June 30 2011). This is why the perceptions of sign users, as registered in the post-test questionnaires, should allow for a better understanding of a sign's status (symbolic vs. iconic / indexical).

The post-test questionnaires provide some really useful insights into the mechanisms driving the interacting pairs' gestural communication and their alignment to a

shared code. Participants in both conditions were asked if while playing they followed any particular strategy in gesturing the target meanings. Participants who started the game as directors based their gestures on salient features of the objects; on how the objects on the cards are used or look like (shape), and also in some cases on some other properties related to producing sound (e.g. speaker, alarm clock, speaker) or the emotions they cause (e.g. lemon was communicated in the majority of cases by a facial expression denoting sourness). These comments underline iconicity's key role in establishing communication throughout the interaction, but also that gestures preserved their non-arbitrary status throughout the experiment.

Participants also note that, after the second round has been played, they had tried to be consistent with the gestures of their partners. Participants who started the game as matchers mostly noted that the strategy they followed was "following the strategy of their partner" (i.e. pantomiming salient features of an object) as well as copying ('imitation') the gestures of their partners since they were the ones gesturing first. All participants noted that they tried to employ gestures similar to the ones of their partner in order to communicate faster and most efficiently (they highlight 'consistency' as a decisive factor for successful communication – although they did not actually know if they were successful). Some of them commented on the slight difficulty in gesturing similar items, albeit this was not considered enough to hinder successful communication since they were able to gesture out features that were "unique" for each object.

Participants, again in both conditions, were also asked if they had noticed their partners following a particular strategy when gesturing the concepts and how they knew to what object the gestures referred to. In a nutshell, what they noticed was that both they and their partners were following a similar strategy. Participants starting as matchers noted that their partners based their gestures mostly on how the objects are used or look like. They used the experience of previous rounds to know which gestures were used for what, and, if communication was delayed in a previous round, they altered and/ or elaborated them. Participants who started as directors noted that the strategy their partner followed in gesturing in the second round was pantomiming salient characteristics of an object related to appearance or usage, using gestures, however, which were similar to what they themselves

created in the first round. These answers reveal that gestures probably remained iconic for the people who created them, and thus held a non-arbitrary relation to their referents. Participants also mention that after the second round they also copied the modifications added by their partner if that allowed them to communicate faster and easier.

The questionnaires verify what the aforementioned statistical analyses, and observations coming from the video-recordings, suggest; iconicity is highly important for establishing communication in the first place. Once communication is established, interaction – involving imitation and some kind of indirect feedback – consolidates faster and more efficient communication. Apart from that, the consistency in using similar gestures is what allowed individuals within pairs to converge upon a locally shared code.

Picture 11: Gesture for 'wine bottle' across different pairs

Picture 12: Gesture for 'alarm clock' across different pairs

Picture 13: Gesture for 'champagne' across different pairs

Picture 14: Gesture for 'water bottle' across different pairs

Examining the data from the final ('sound only') round of the gesture-and-sound condition, it becomes apparent that in some instances a pair might have used certain sounds consistently with particular objects throughout the first 4 rounds, but for specific reasons failed to communicate in the final round. These reasons are mainly linked to the fact that in communication both interlocutors must align to the same code in order to understand each other; if I only speak Greek and you only speak Chinese, no matter how good we are in our respective languages we will still be unable to communicate (this of course implies only the use of a spoken language and no other means of communication such as gestures, which people use when found in similar situations). This seems to have been the case in several instances in the interacting gesture-and-sound condition, but it was extremely prominent in a particular pair. In this pair (participants B3 & B4), the two participants used specific sounds consistently with specific gestures, but each of them had a sound of his/her own. This did not affect their performance in the first 4 rounds of the game since their identification accuracy performance follows the same trend presented in the graphs we examined above (R1: 15/24; R2: 20/24; R3: 23/24, R4:22/24). Hence, gesture in this case had the role of an "intermediary" allowing the pair to communicate successfully. Nevertheless, when confronted with the possibility of communication only by using sounds, this pair failed completely (R5: 0/24). To exemplify this even further, when communicating the concept of 'lemon' in Rounds 1-4, participant B3 used along with his/her gesture the sound 'isp' (in both Round 1 & 3), whereas participant B4 used the sounds 'skroth' (Round 2) and 'fruhv' (Round 4). In Round 5, both participants had the opportunity to play this object as directors/matchers. Participant B3 used the sound 'isp' to communicate 'lemon', but participant B4 mistakenly interpreted this as 'staples' since this was the sound he used for this particular object. On the other hand, participant B4 used the sound 'fruhv' but B3

interpreted it as 'champagne' (this was probably a random guess since s/he had never paired this sound with any of the concepts in previous rounds).

The overall performance in these rounds reveals the possibility that gesture can indeed function a bootstrapping device between meaning and sound. Despite the variability in performance it is important to note that 3 of the 7 pairs managed to attach a sound to more than half of the target objects. The importance of these results attains an even greater status if we take into account the relatively limited time of interaction, but most importantly, the lack of feedback; the reciprocity of the interaction in combination with the communicative power of gesture, deriving from its inherent iconicity, is sufficient and efficient for boosting mutual intelligibility in the vocal-auditory channel. In spite of the novel words being unnecessary for communication, participants started using them communicatively, by virtue of grounding via gesture. Contrary to Garrod et al's (2007) study, in which it is debatable if symbols actually emerge from icons, this experiment shows that symbols *can* arise in a situation involving a multimodal (gesture & sound) and semiotically mixed (non-arbitrary & arbitrary) communication system. In the gesture-and-sound condition, even though the novel words were unnecessary for achieving successful communication, they came to carry meaning in an arbitrary way simply by being there and repeatedly being associated with meaningful iconic gestures. Hence, gesture was what initially allowed pairs to converge on a locally shared communication system and to eventually expand this system to a different modality using arbitrary sound-meaning pairings to communicate target meanings.

It could be argued that in the majority of cases if we allowed the pairs to interact for some additional time they would have been able to show improved results; if they have started giving meaning to some of the words it would be surprising if gradually they could not give meaning to all. Of course this is in some ways a half-full, half-empty situation. Even though bootstrapping is possible, we cannot ignore the fact that identification accuracy decreases approximately to 50% when gesture is taken away as an option (Figure 6). It would require a relative amount of time and effort to bootstrap a much larger lexicon instead of the relatively small one I tested in the experiment (there is no correlation between

a pair's identification accuracy success in the gesture-and-sound rounds (Round 1 - 4) and its success in the sound only round (Round 5)[19]).

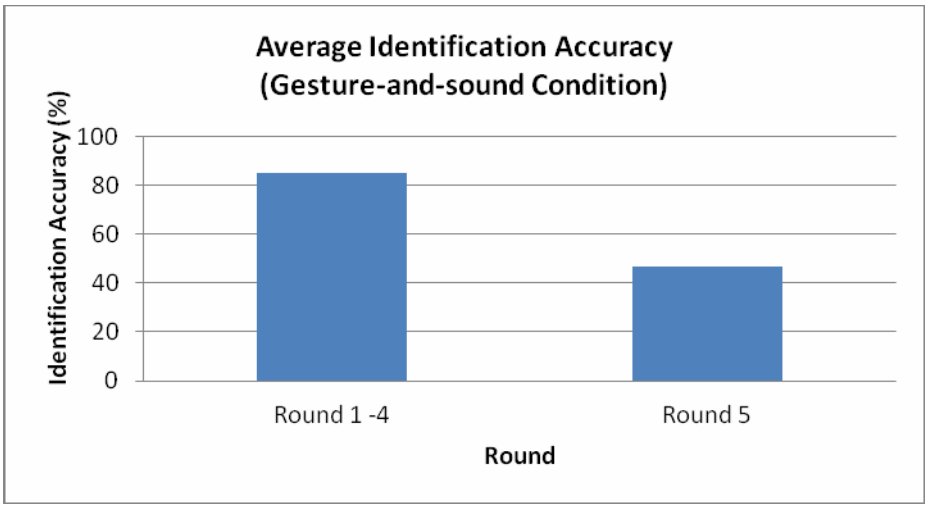


Figure 6: Identification accuracy in the first 4 Round of the gesture-and-sound condition compared to the identification accuracy in the 5th Round of the same condition where gesture is restricted (sound-only)

The post-test questionnaires also provide useful insights into the strategies participants used in choosing and employing the novel words. In asking them whether they chose the word they used randomly most participants answered that they did so for some items while for others they chose them based on onomatopoeia and spelling similarities to the actual words denoting the items. As far as onomatopoetic clues that might have facilitated the participants, it is noteworthy that everyone appears to have a different opinion of which words were onomatopoetic; thus, these kinds of observations are highly subjective. For instance, two pairs have successfully used 'splow' to communicate champagne in the sound-only round. Although, in retrospect this sound might be similar to a splashing sound, no other pair was able to use the same strategy to communicate this object. On the contrary, the word 'fruhv' was used by another two pairs, whereas 'krif' and 'isp' were used by another pair respectively. Two participants of different pairs noted the sound 'fruhv' as bearing similarity to the word 'fruit'. Despite the validity of this observation, only one of them was able to use this sound to successfully communicate the concept of 'lemon'. The other participant was not able to do so, something implying that her partner did not share her view. In fact these were the only instances where this sound was coupled with one of the three fruits (lemon, orange, pepper), whereas six other sounds were

[19] sig. prop. (2-tailed) > .01 (see APPENDIX E for details).

used to communicate them successfully in the sound-only round. One participant who in the 'sound-only' round was able to communicate 'tape recorder' through the sound 'broiced' said he chose this sound due to its similarity to 'voiced'. However, nobody else was able to communicate the same object through this sound. The same applies for two pairs who used 'oldge' to communicate 'orange'.

Overall, it is rather unlikely that onomatopoeia or spelling similarity have confounded the results. This is because, in order to put them in a good use, participants had to overcome three communication barriers inherent in their task: (1) the fact that within a pair both participants had to share perspectives in choosing sounds; (2) that they could not exchange feedback in order to make their partners aware of their perspectives; and (3) that the item list contained easily confusable items (e.g. three fruits instead of one). Even if some words may have been perceived as non-arbitrary to a limited degree, it does not follow that they were not perceived this way by others. If only one member within a pair interpreted a particular sound as non-arbitrary, this was not enough to facilitate communication in the vocal channel. It is like looking at the clouds in the sky and making judgements about what they resemble – you might see an elephant but I might still see a cloud. Even if we consider a scenario in which both of the participants interpret a word non-arbitrarily but disagree on the meaning expressed by the word, the two judgments cancel each other out. Thus, the first person who will adopt her partner's sound will interpret it in an arbitrary way, since she will be unfamiliar with the reasons underlying her partner's choice.

This argument is verified by the fact that the participants' answers to the aforementioned question are slightly contradictory to their answers to "How they knew what the words meant?". This is because, with the exception of a single pair, all participants responded that the decisive factor for knowing what a word meant was: (a) using the word consistently along a particular gesture across rounds; and (b) trying to remember and preserve the word their partners had used in the previous round. In spite of whatever "clues" participants tried to find in the sound and spelling of words, it was the interaction with their partners that led to the establishment of shared meanings for words.

It becomes evident that a shared lexicon in the vocal-auditory channel can indeed be established through repeated and reciprocal association of sounds to non-arbitrary bodily

actions. Communication is a joint activity and iconicity can indeed lie at the roots of the evolution of communication systems; iconicity in this experiment was essential for the construction of a symbolic communication system (a lexicon). As it was mentioned above, participants commented on the fact that in order to know what a gesture meant they based their judgements on an object's salient characteristics instantiated in their partners' gestures. Consequently, on the one hand, we have the gestures, which the participants interpreted based on their similarity to their referents, and on the other, the novel words which the participants interpreted by means of repeated association to a particular gesture. If that gesture-sound match was consistent, i.e. if they copied what their partner produced in a previous round, then a meaning was successfully attached to the sound. Even though (some of) the iconic features of gestures have been retained, forms in the vocal channel took on a symbolic form. This is in some ways similar to what we find in language, with multimodality and semiotically mixed forms (see Brown in prep. a).

Convergence on a shared code in the gestural modality is thus proven vital for successfully converging on a symbolic code in the oral modality. The reason feedback is not required to give meaning to meaningless words is because with gestures the feedback is in the gestures themselves. This is because the iconicity inherent in pantomime provides insights into a person's intended meaning. Thus, even if the director cannot directly inform her partner about the target meaning's identity, the iconic features of her gesture can (indirectly) provide sufficient information about what the target meaning might be. Gestures make possible what arbitrary sounds alone could not achieve by themselves in a pure 'sound-only' condition; as it was previously said, (arbitrary) sounds are iconically weaker than gestures since they cannot resemble their referents as efficiently as gestures can (Brown in prep. a). Thus, it would have been practically impossible for participants to communicate by means of arbitrary sounds alone.

The last question included in the questionnaires asked the participants whether they used the sounds before or after gesturing and what was the reason underlying their choice. There does not seem to be a ubiquitous strategy participants followed as far as positioning the sounds is concerned; some say they produced the words before, while others say they did so simultaneously or after. Nevertheless, by examining the videos it becomes evident

that the participants who said they pronounced the word after had the wrong impression[20]. All participants produced the words either before or along gesturing (some had a fixed strategy – only before – while others had a mixed one – some words before & some simultaneously). What is remarkable is that the participants who produced the words simultaneously with their gestures sometimes altered the tone of their voice or prolonged the pronunciation of the sound to make it fit with the gesture. For example, a pair employed a yawning gesture to communicate the concept of an 'alarm clock' and produced the sound 'yemf' simultaneously (see Picture 15). The sound's duration matched the one of the gesture, something that was established in both pair members and gradually became a unified, multimodal and semiotically mixed (symbolic & non-arbitrary) sign for 'alarm clock'. Unsurprisingly, 'yemf' was successfully used to communicate in the sound-only round. It would be interesting to examine this factor alone with a larger sample and for a longer period of time (more rounds) to check whether simultaneous and integrated production can facilitate the association of sounds to gestures, i.e. if the synchronic use of gesture and sound counteracts memory constraints and helps conventionalization of arbitrary forms.

Picture 15: Gesture for 'alarm clock'. In the first 4 snapshots (all taken from Round 1) the director produces the sound 'yemf' simultaneously with a yawning gesture. In the 5th snapshot we have the simultaneous gesture and sound production as it was established in Round 4

[20] With the exception of a limited number of occasions in which participants forgot to produce a sound and they had to be reminded in order to do so.

8. Concluding remarks and Future directions

The experimental work just described examined the evolution of a gestural communication system and its usefulness for establishing a communication system in the vocal-auditory channel. It was attempted to demonstrate that gestures (pantomime) *can* serve as a bootstrapping device between meanings and arbitrary sounds, if the latter are repeatedly associated with iconic gestural signs in an environment involving role reversal imitation / the context of reciprocal pair-wise interactions. We do not wish to claim that the evolution of a multimodal communication system as this was described in my experiment is firmly linked to what actually happened at the early stages of language evolution. What was shown is that arbitrary conventions can arise as the result of having a multi-modal (gesture + sound) and semiotically mixed (non-arbitrary + arbitrary signs) communication system; the iconically more powerful gestures can ground communication, while the iconically weaker vocalizations can attain meaning as a result of repeated interaction and consistent association with meaningful gestures. This is to highlight the fact that iconicity, encapsulated in gestures, might have been an indispensable ingredient of the roots out of which human communication systems sprang. Arbitrary and conventional oral communication could have, at least, initially overlapped with more naturally meaningful gestures (pantomimes and points) which concatenated meanings with formerly meaningless vocalizations.

The central experimental hypothesis revolving around the possibility of creating a shared communication system (a lexicon) based on gestures through a "grounding process" (Clark & Brennan, 1991), which can in turn be used to lend meaning to previously meaningless sound strings, was verified. The majority of individual hypotheses concerning the development of signs in the two experimental conditions were also verified. Gestures in both the gesture-only and the gesture-and-sound condition became increasingly shorter. Similarly to Garrod at al.'s (2007) experiment, interacting dyads in both conditions become increasingly better at identifying each other's signals. As the interacting participants alternated roles (director/ matcher) their gestures progressively converged with those of their partners. This happens because this kind of local interactions produce accurate local

sign systems that are shared within the interacting dyads. The key in the simplification and refinement of gestures lies in the reciprocity of the interactions (participants acting both as directors and matchers).

Most importantly, even in the absence of feedback, "simplification-by-grounding" (Garrod et al. 2007, p. 966) is still attainable. At the early stages of communication, iconicity is central in establishing common ground. Iconicity sanctions the online and constant renegotiation of meaning, before the ascertainment of a shared set of signs. With respect to the gesture-and-sound condition, participants were indeed capable of aligning upon a locally shared multi-modal and semiotically mixed communication system using both gestures (pantomime) and the available novel words in order to communicate. Gestures served, at a significant degree, as a bootstrapping device between target meanings and arbitrary sounds allowing the interacting pairs to eventually use arbitrary sound-meaning pairings to communicate the target meanings. Thus, it could be argued that gestures allowed (some of) the new words to attain meaning and become conventionalized. Despite the lack of feedback, reciprocal interaction in combination with the iconicity inherent in gestures is sufficient and efficient for boosting mutual intelligibility in the vocal-auditory channel.

Against our initial expectations, gestures in the gesture-and-sound condition did not become shorter than those in the gesture-only condition. On the contrary, they remained longer. Apart from that, the identification accuracy in the gesture-and-sound condition was surprisingly worse than in the gesture-only condition. This result could be attributed to the fact that participants in the gesture-and-sound condition found it hard to remember which sound corresponded to which meaning/ gesture. This might have hindered both the identification, but also the ease of gesturing the target concepts, since participants might have dealt with some kind of extra cognitive/ memory load. Sounds remained supplementary throughout the experiment, whereas gestures were preserved as the prominent medium of communication. This is quite interesting since it indicates that using novel arbitrary signs can in some way hinder communication - it creates tension - which suggests that a transition to/adoption of new symbols is not trivial and also that there needs to be a rich, effective means of communication (in this case gestures) that can maintain

comprehension/grounding so that arbitrary signs can become established over time (E. Brown, personal communication, June 30 2011).

Although gestures lost some of their initial iconic features due to their simplification and refinement, it cannot be argued that they have transformed into symbols. It might indeed be the case that the information starts to shift from the signs structure to the user's knowledge of the sign's previous use (Garrod et al. 2007, p. 964) since the participants' comments seem to suggest that what they gradually paid attention to was the copying their partners' gestures. Having that said, the similarity in the gestures of non-interacting participants reveals the possibility that gestures can remain iconic and at the same time be somewhat shared between people who have never interacted before (having high transparency). In order to be certain if signs were used arbitrarily (i.e. as a symbol) or non-arbitrarily (i.e. as an icon or index), the experiment examined how participants perceived their signs' form in the context of this communication game (ibid.). The post-test questionnaires reveal that gestures probably remained iconic for the people who created them, and thus held a non-arbitrary relation to their referents.

Vocalizations, being iconically weaker since they cannot resemble their referents as efficiently as gestures can (Brown in prep. a), are not as good as gestures to establish communication when a shared code is absent. It still remains to be seen if a 'sound-only' condition in the absence of feedback will result in unsuccessful communication. What can be also tested is how participants would perform in a gesture-and-sound condition in which, however, they would have to generate the sounds themselves (N. Fay, personal communication, June 26, 2011). In this kind of condition, participants would probably employ a combination of non-arbitrary gestures and non-arbitrary sounds (onomatopoetic words) in order to communicate (ibid.). The participants' tendency to look for onomatopoeia in the words they were given (although they were told that the words were novel and arbitrary), plus that some of them integrated them with their gestures, implies participants in a future experiment may perform better under a non-arbitrary sound condition. These assumptions can be tested in a future experiment which will adapt the same design with the discussed study. In order to have more direct comparisons with Garrod and colleagues' (2007) study both a 'Single Director with Feedback' and a 'Double Director with Feedback'

conditions can be examined across the abovementioned 'Gesture-only' and 'Gesture-and-sound' conditions.

The situation described in the experiment resembles to a great extent what we find in homesign systems. Based on these gestures, homesigners are able to create some basic systems which incorporate an open-ended lexicon, a simple morphological system (inflectional and derivational) and a simple syntactic structure (Goldin-Meadow& Mylander 1990). However, these children are confined in a situation where they have to preserve the transparency of their symbols since this is fundamental for their decipherability by outsiders (Morford 1996). The lack of a community to share a communicative code is fundamental in explaining the inability of homesigns to develop into full-blown languages. The newly emergent Nicaraguan Sign Language provides substantial evidence for this argument. When brought together in a shared environment, deaf children with different homesign systems started creating a common language (a 'pidgin'), systematizing the input they got from each other. As younger children were exposed to this more systematized system, a new language (a 'creole') was created which gradually became increasingly complex with more conventionalized and less transparent symbols (Kegl et al. 1999).

It would be really interesting to examine the evolution of a gestural communication system in an experimental design involving a more dynamic interaction between the participants. Thus, instead of interacting dyads (local communication systems) a future experiment can involve communities of participants trying to establish global communication systems. Under this kind of conditions, it will be interesting to examine whether gestural signs will remain iconic, or if they will start to attain a more symbolic status as we shift from pair-wise to population-level interactions. In a similar experiment again with graphical representations, Fay and colleagues (2008) found that community-evolved signs, in spite of having equal visual complexity to locally-evolved signs, retained a relative amount of residual iconicity. This helped the improvement of individual learning and the efficient translation from sign to meaning (ibid.). It has become clear that gesture should be considered an indispensable component of human communication. The experimental paradigm just described, based on playing charades, was used to argue for the substantial contribution of gestures in the creation of human communication systems. The

same paradigm can be exploited to investigate even further the effect non-arbitrary gestural representations have on symbolic communication and more specifically the mechanisms driving the transformation of icons into symbols.

References

Arbib, M. A. (2005). Interweaving protosign and protospeech: Further developments beyond the mirror. *Interaction Studies,* 6 (2), 145-171.

Armstrong, D. F., Stokoe W. C., Wilcox, S. E. (1994). Signs of the origin of syntax. *Current Anthropology*, 35(4), 349-368.

Barakat, R. A. (1969). Gesture systems. *Keystone Folklore Quarterly*, 14, 105 – 121.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480normative photos of objects to be used as visual stimuli in cognitive research. *PLoS One*, 5 (5), 1-13.

Browman, C. P., & Goldstein, L. (1991). Gestural structures: distinctiveness, phonological processes, and historical change. In I. G. Mattingly and M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory of Speech Perception* (pp. 313-335).Hillsdale, NJ: Erlbaum.

Botha, R. (2007). On homesign system as a potential window on language evolution. *Language & Communication*, 27, 41-53.

Brown, E. (in preparation a). A multimodal perspective on the emergence of symbolic communication.

Brown, E. (in preparation b). Reconciling theory and practice in experimental semiotics.

Brun, T. (1969). *The International Dictionary of Sign Language*. London: Wolfe.

Clark, H. H, & Brennan, S. A. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127 – 149). Washington, DC: American Psychological Association.

Corballis, M. C. (2003). From hand to mouth: The gestural origins of language. In M.H. Christiansen & S. Kirby (Eds.), *Language Evolution* (pp. 201-218). Oxford University Press.

Corballis, M. C. (2010). Mirror neurons and the evolution of language. *Brain & Language*, 112, 25-35.

de Jorio, A. (1832). *La Mimica Degli Antichi Investigate nel Gestire Napoletano*. Naples: Fibreno.

Fay, N., Garrod, S., & Roberts, L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 363, 3553-3561.

Garrod, S., Fay, N., Lee, J, Oberlander, J, & MacLeod, T. (2007). Foundations of Representation: Where might graphical representation symbol systems come from?. *Cognitive Science*, 31, 961-987.

Goldin-Meadow, S. (2005). The resilience of language: What gesture creation in deaf children can tell us about how all children learn language. New York: Psychology Press.

Goldin-Meadow, S. & Mylander, C. (1990). Beyond the input given: The child's role in the acquisition of language [Electronic version]. *Language*, 66 (2), 323-355.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally [Electronic version]. *PNAS*, 105 (27), 9163-9168.

Iverson, J., Caprici, O., & Caselli, M. C. (1994). From communication to language in two modalities. [Electronic version]. *Cognitive Development*, 9, 23-43.

Iverson, J., & Goldin-Meadow, S. (2005). Gesture paves the way for language development [Electronic version]. *Psychological Science*, 16, 367-373.

Hurford, J. R. (2004). Language beyond our grasp. In K. Oller, U. Griebel, & K. Plunkett (Eds.), *Evolution of Communication Systems: A Comparative Approach* (pp. 297–313). Cambridge University Press, Cambridge, MA.

Kegl, J, Senghas, A., & Coppola M. (1999). Creation through contact: Sign language emergence and sign language change in Nicaragua. In M. DeGraff (Ed.), *Language creation and language change: Creolization, diachrony, and development* (pp. 179-237). Cambridge, MA: MIT Press.

Kendon, A. (1981). Geography of gesture. *Semiotica*, 37, 129-163.

Kendon A., (1988). *Sign Languages of Aboriginal Australia: Cultural Semiotic and Communicative Perspectives*. Cambridge: Cambridge University Press.

Kendon, A. (1997). Getsure. *Annual Reviews of Anthropology*, 26, 109-128.

Landsberg, M. E. (1980). The icon in semiotic theory. *Current Anthropology*, 21 (1), 93-95.

Maybery, R. I., & Nicoladis, E. (2000). Gesture reflects language development [Electronic version]. *Current Directions in Psychological Science*, 9 (6), 192-196.

McNeill, D. (1992). *Hand and Mind: What Gesture Reveal About Thought.* Chicago: Chicago University Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago: Chicago University Press.

McNeill, D., Duncan, S., Cole, J., Gallagher, S., & Bertenthal, B. (2008). Growth points from the very beginning. *Interaction Studies*, 9, 117–132.

Morford, J. P. (1996). Insights to language from the study of gesture: A review of research on the gestural communication of non-signing deaf people. *Language & Communication*, 16 (2), 165-178.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. New York: Freeman.

Parrill, F. (2008). Form, meaning, and convention: A comparison of a metaphoric gesture with an emblem. In A. Cienki and C. Müller (Eds.), *Metaphor and Gesture* (pp. 195-217). Amsterdam: John Benjamins.

Peirce, C. S. (1931–1958). *Collected papers of Charles Sanders Peirce.* Vols. 1–8. Charles Hathorne, Paul Weiss, & Arthur Burks (Eds.). Cambridge, MA: Harvard University Press.

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (ed.), *Metarepresentations: An Interdisciplinary Perspective* (pp.117-137). Oxford: OUP.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuo-motor structure. In D. A. Allport (Ed.), *Language Perception and Production* (pp. 67-84). New York: Academic Press.

Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.

Zlatev, J., Persson T., Gärdenfors, P. (2005). Bodily mimesis as the "missing link" in human cognitive evolution. *Lund University Cognitive Studies*, 121, 1-40.

Zlatev, J. (2008). From proto-mimesis to language: Evidence from primatology and social neuroscience. *Journal of Physiology-Paris*, 102 (1-3), 137-151.

Appendix A: Experimental Materials

Picture 1: Example sheet with all items for Participant A

Picture 2: Example sheet with all items for Participant B

krif

gwuk

yemf

vilb

yeents

altch

splow

skroth

snen

fruhv

kreldge

djorp

plect

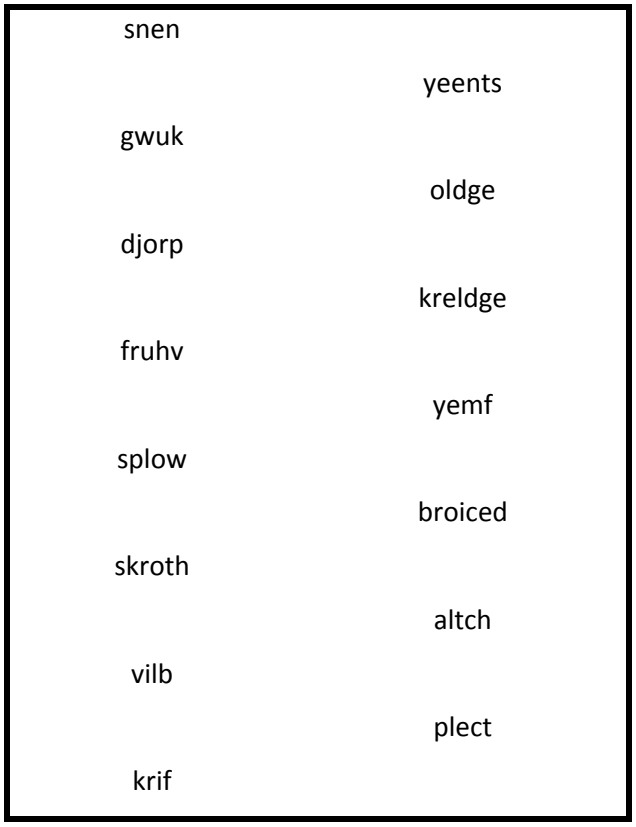
oldge

isp

broiced

Picture 3: Sheet with all non-words for Participant A

isp



Picture 4: Sheet with all non-words for Participant B

isp

snain

fruhv

kreldge

djorp

vilb

yeents

altch

splow

krif

gwuk

yemf

broiced

skroth

oldge

plect

Table 2: List of English non-words

Questionnaire

PARTICIPANT ID : ..... DATE:.....

(Questions 3,4, & 5 are applicable only for participants in the second experimental condition)

1. **Did you follow any particular strategy when gesturing the concepts?**

   .....
   .....
   .....
   .....

2. **Did you notice your partner following a particular strategy when gesturing the concepts (i.e. How did you know to what the gestures referred to)?**

   .....
   .....
   .....
   .....

3. **Did you choose the "words" you used randomly?**

.....

.....

.....

.....

4. **How did you know what the "words" meant?**

.....

.....

.....

.....

5. **Did you produce the sounds given to you before or after your gestures? Why?**

.....

.....

.....

.....

Game Instructions (Gesture-only Condition)

For Directors:

1. Take a card from the pile that is situated on the table
2. Carefully turn the card towards the video camera revealing its content
3. Do not reveal the card's content to your partner
4. Gesture with any body part you want to communicate the object depicted on the card
5. Do not use any spoken language or any other kind of vocalization
6. Once your partner has made her/ his guess, reveal your card to him/ her
7. Discard the card you have just used
8. Return to 1

For Matchers:

1. Observe the gestures performed by your partner
2. Try to guess which of the objects depicted in the cards in front of you your partner tries to communicate through gesture

3. Once you think you have found the answer call your partner to "STOP" and pick up the relevant card
4. Reveal your card to your partner and to the camera
5. Do not speak or gesture towards your partner in order to provide any other kind of feedback

Game Instructions (Gesture-and-sound Condition)

For Directors:

1. Take a card from the pile that is situated on the table
2. Carefully turn the card towards the video camera revealing its content
3. Do not reveal the card's content to your partner
4. Select a sound from the list of "words" provided to you to represent the object in the card
5. Use any body part you want **AND** the "word" you selected to communicate to your partner the object depicted on the card (using both is mandatory)
6. Do not use any spoken language or any other kind of vocalization apart from the "word" you selected
7. Once your partner has made her/ his guess, reveal your card to him/ her
8. Discard the card you have just used
9. Return to 1

For Matchers:

1. Observe the communicative behaviour of your partner
2. Try to guess which of the objects depicted in the cards in front of you your partner tries to communicate through the sound and gesture s/he produces
3. Once you think you have found the answer call your partner to "STOP" and pick up the relevant card
4. Reveal your card to your partner and to the camera
5. Do not speak or gesture towards your partner in order to provide any other kind of feedback

ANSWER SHEET (1)

TRIAL	GAME 1	GAME 2	GAME 3	GAME 4
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				

Table 3: Answer Sheet for Gesture-Only Condition

ANSWER SHEET (2)

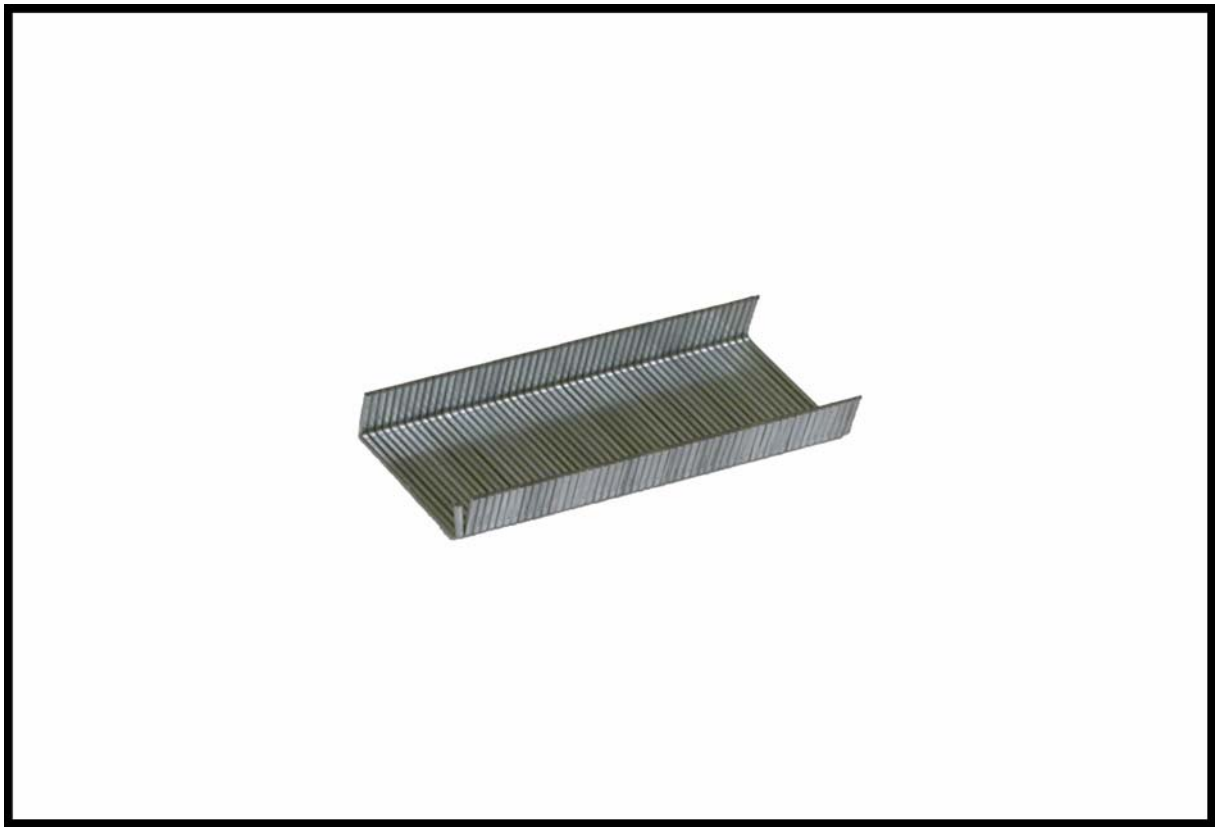
TRIAL	GAME 1	GAME 2	GAME 3	GAME 4	GAME 5
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

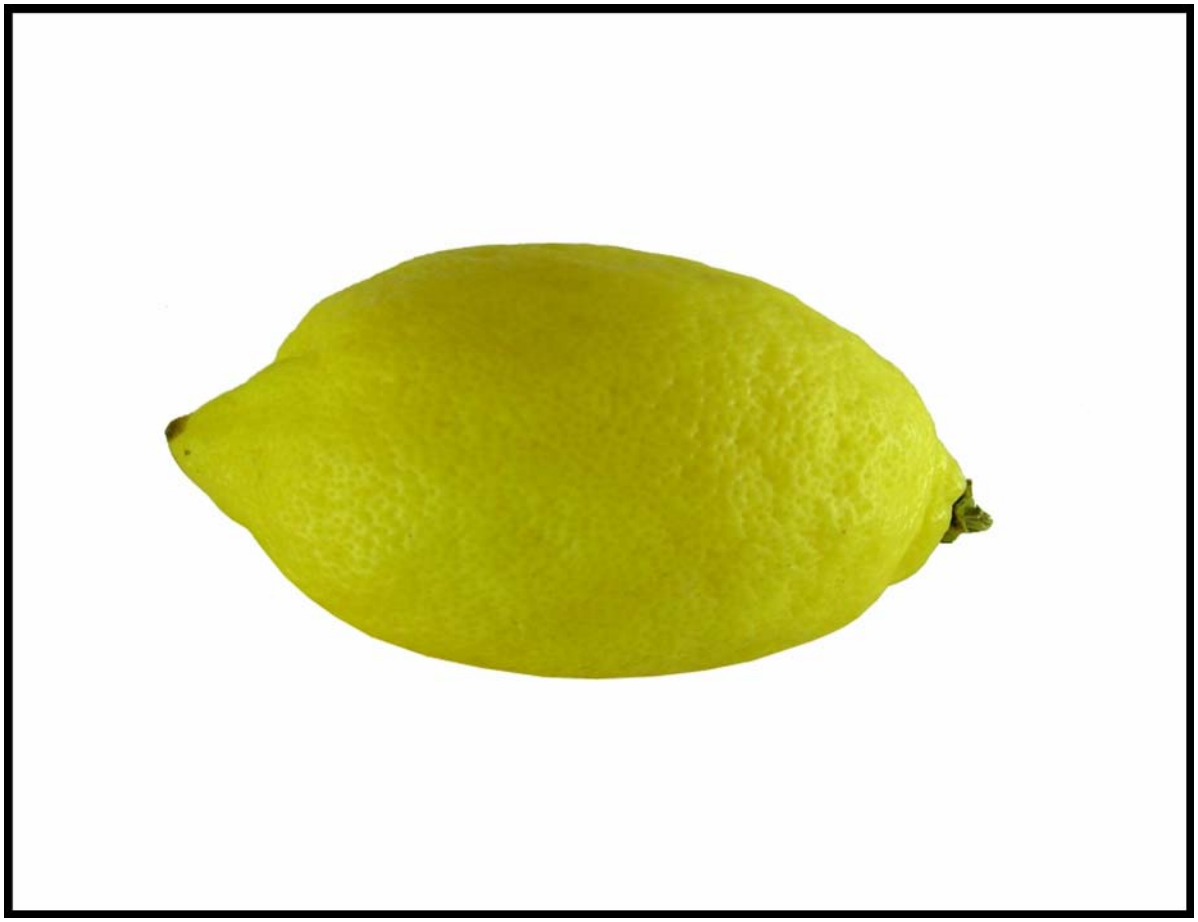
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					

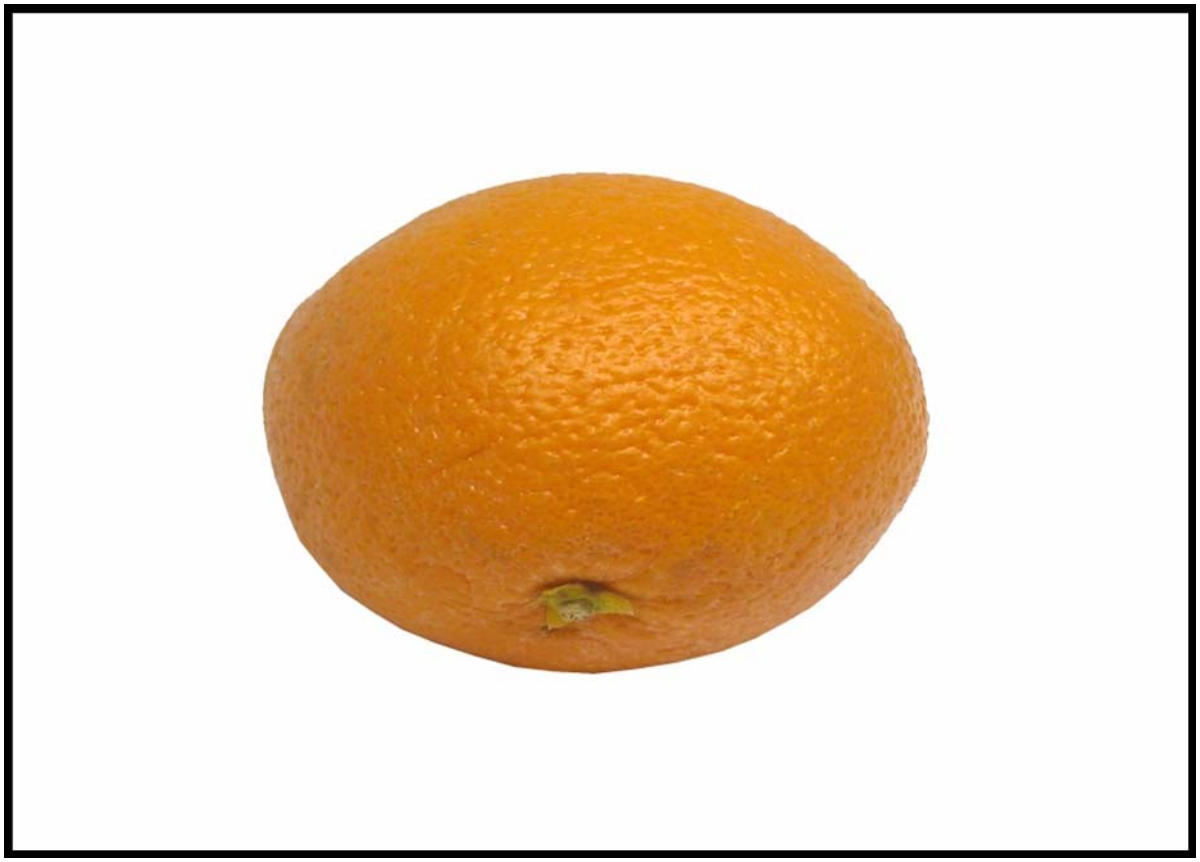
Table 4: Answer Sheet for Gesture-and-sound Condition


H
M
WAKE/
ALM OFF
PM
AUTO


LABYRINTH
LABYRINTH







Appendix B: Round 5 Identification Accuracy Statistics

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Correct Answers in Round 5	7	11.2857	7.18132	2.71429

One-Sample Test

	Test Value = 2					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Correct Answers in Round 5	3.421	6	.014	9.28571	2.6441	15.9273

One-Sample Test

	Test Value = 4					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Correct Answers in Round 5	2.684	6	.036	7.28571	.6441	13.9273

One-Sample Test

	Test Value = 5					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Correct Answers in Round 5	2.316	6	.060	6.28571	-.3559	12.9273

Appendix C: Identification Accuracy

By-subject Analysis

Mauchly's Test of Sphericity (By-subject analysis)

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
round	.663	4.413	5	.493
manipulability	1.000	.000	0	.
round * manipulability	.133	21.651	5	.001

Tests of Within-Subjects Effects (By-subject analysis)

Source		df	F	Sig.
round	Sphericity Assumed	3	7.826	.000
	Huynh-Feldt	3.000	7.826	.000
round * condition	Sphericity Assumed	3	.715	.549
	Huynh-Feldt	3.000	.715	.549
Error(round)	Sphericity Assumed	36		
	Huynh-Feldt	36.000		
manipulability	Sphericity Assumed	1	2.729	.124
	Huynh-Feldt	1.000	2.729	.124
manipulability * condition	Sphericity Assumed	1	3.183	.100
	Huynh-Feldt	1.000	3.183	.100
Error(manipulability)	Sphericity Assumed	12		
	Huynh-Feldt	12.000		
round * manipulability	Sphericity Assumed	3	.363	.780
	Huynh-Feldt	1.770	.363	.674
round * manipulability * condition	Sphericity Assumed	3	.430	.733
	Huynh-Feldt	1.770	.430	.632
Error(round*manipulability)	Sphericity Assumed	36		
	Huynh-Feldt	21.235		

Tests of Between-Subjects Effects (By-subject analysis)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	13050.723	1	13050.723	1400.525	.000
condition	58.580	1	58.580	6.286	.028
Error	111.821	12	9.318		

Mauchly's Test of Sphericity (By-subject analysis)

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
round	.663	4.413	5	.493
manipulability	1.000	.000	0	.
round * manipulability	.133	21.651	5	.001

Tests of Within-Subjects Effects (By-subject analysis)

Source		df	F	Sig.
round	Sphericity Assumed	3	7.826	.000
	Huynh-Feldt	3.000	7.826	.000
round * condition	Sphericity Assumed	3	.715	.549
	Huynh-Feldt	3.000	.715	.549
Error(round)	Sphericity Assumed	36		
	Huynh-Feldt	36.000		
manipulability	Sphericity Assumed	1	2.729	.124
	Huynh-Feldt	1.000	2.729	.124
manipulability * condition	Sphericity Assumed	1	3.183	.100
	Huynh-Feldt	1.000	3.183	.100
Error(manipulability)	Sphericity Assumed	12		
	Huynh-Feldt	12.000		
round * manipulability	Sphericity Assumed	3	.363	.780
	Huynh-Feldt	1.770	.363	.674
round * manipulability * condition	Sphericity Assumed	3	.430	.733
	Huynh-Feldt	1.770	.430	.632
	Sphericity Assumed	36		

Pairwise Comparisons (Condition) (By-subject analysis)

(I) condition	(J) condition	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
gesture only	gesture and sound	1.446*	.577	.028	.189	2.703

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Univariate Tests (Condition) (By-subject analysis)

	Sum of Squares	df	Mean Square	F	Sig.
Contrast	7.323	1	7.323	6.286	.028
Error	13.978	12	1.165		

Pairwise Comparisons (Round) (By-subject analysis)

(I) round	(J) round	Mean Difference (I-J)	Std. Error	Sig.[a]
1	2	-.786	.323	.191
	3	-1.250*	.343	.020
	4	-1.000*	.271	.019
2	1	.786	.323	.191
	3	-.464	.223	.357
	4	-.214	.215	1.000
3	1	1.250*	.343	.020
	2	.464	.223	.357
	4	.250	.237	1.000
4	1	1.000*	.271	.019
	2	.214	.215	1.000
	3	-.250	.237	1.000

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

*. The mean difference is significant at the .05 level.

Pairwise Comparisons (Manipulability) (By-subject analysis)

(I) manipulability	(J) manipulability	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
Low Manipulability	High Manipulability	-.446	.270	.124	-1.035	.142

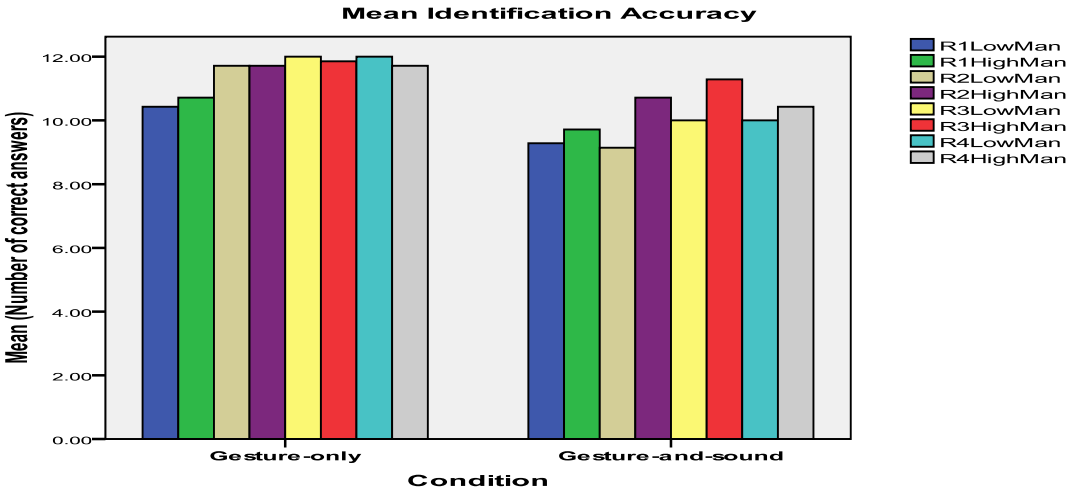
Based on estimated marginal means

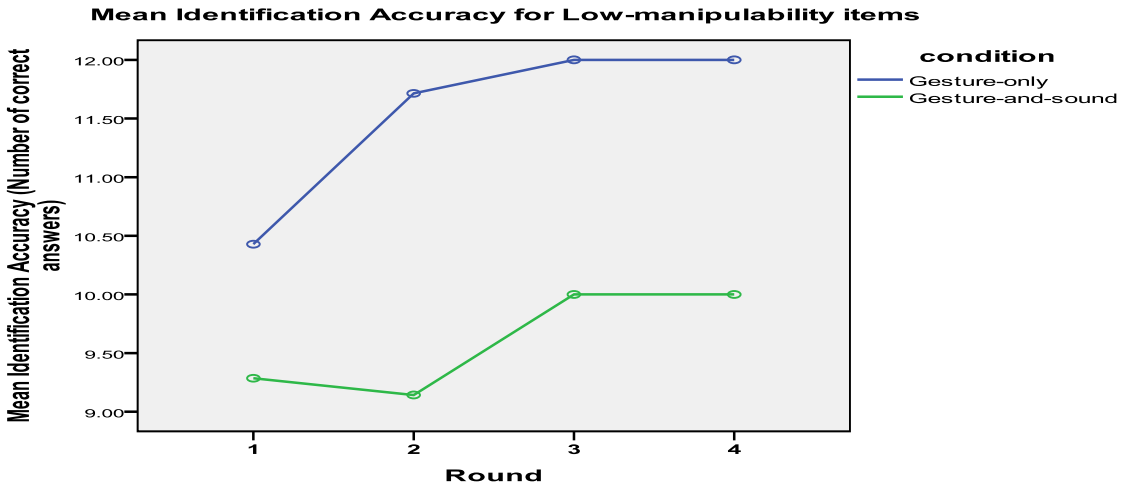
Pairwise Comparisons (Manipulability) (By-subject analysis)

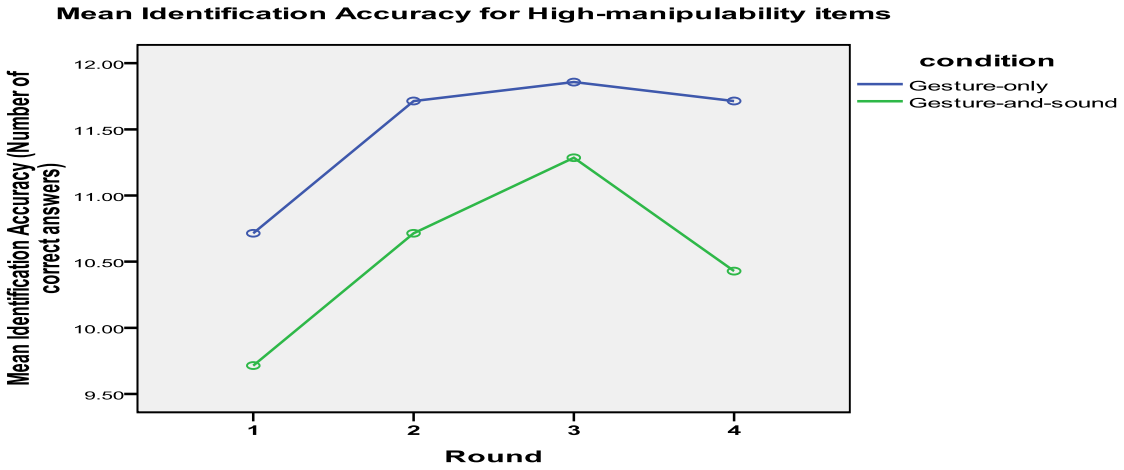
(I) manipulability	(J) manipulability	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
Low Manipulability	High Manipulability	-.446	.270	.124	-1.035	.142

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.







By-item Analysis:

Mauchly's Test of Sphericity (By-item analysis)

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
round	.399	8.006	5	.158
condition	1.000	.000	0	.
round * condition	.339	9.446	5	.094

Mauchly's Test of Sphericity (By-item analysis)

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
round	.399	8.006	5	.158
condition	1.000	.000	0	.
round * condition	.339	9.446	5	.094

Tests of Within-Subjects Effects (By-item analysis)

Source		df	F	Sig.
round	Sphericity Assumed	3	8.948	.000
	Huynh-Feldt	2.886	8.948	.000
round * Manipulability	Sphericity Assumed	3	.713	.552
	Huynh-Feldt	2.886	.713	.547
Error(round)	Sphericity Assumed	30		
	Huynh-Feldt	28.861		
condition	Sphericity Assumed	1	15.467	.003
	Huynh-Feldt	1.000	15.467	.003
condition * Manipulability	Sphericity Assumed	1	1.719	.219
	Huynh-Feldt	1.000	1.719	.219
Error(condition)	Sphericity Assumed	10		
	Huynh-Feldt	10.000		
round * condition	Sphericity Assumed	3	.791	.509
	Huynh-Feldt	2.737	.791	.499
round * condition * Manipulability	Sphericity Assumed	3	.816	.495
	Huynh-Feldt	2.737	.816	.486
Error(round*condition)	Sphericity Assumed	30		
	Huynh-Feldt	27.366		

Pairwise Comparisons (Manipulability) (By-item analysis)

(I) Level of Manipulability	(J) Level of Manipulability	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
Low Manipulability	High Manipulability	-.521	.699	.473	-2.078	1.037
High-Easy	Low-Hard	.521	.699	.473	-1.037	2.078

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

Univariate Tests (Manipulability) (By-item analysis)

	Sum of Squares	df	Mean Square	F	Sig.
Contrast	.814	1	.814	.555	.473
Error	14.659	10	1.466		

Pairwise Comparisons (Rounds) (By-item analysis)

(I) round	(J) round	Mean Difference (I-J)	Std. Error	Sig.[a]
1	2	-.917	.291	.062
	3	-1.458*	.336	.009
	4	-1.167	.391	.082
2	1	.917	.291	.062
	3	-.542	.184	.087
	4	-.250	.312	1.000
3	1	1.458*	.336	.009
	2	.542	.184	.087
	4	.292	.227	1.000
4	1	1.167	.391	.082
	2	.250	.312	1.000
	3	-.292	.227	1.000

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

*. The mean difference is significant at the .05 level.

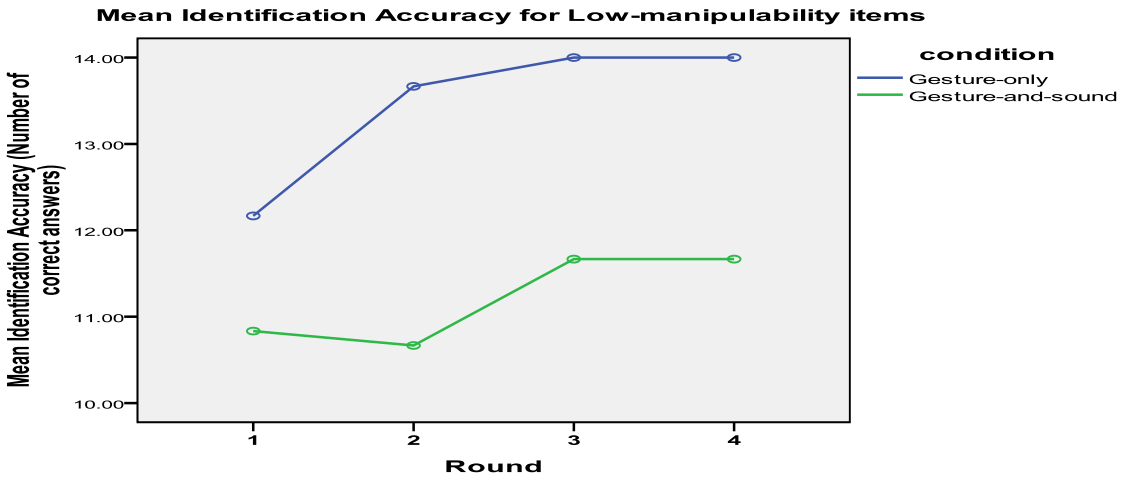
Pairwise Comparisons (Condition) (By-item analysis)

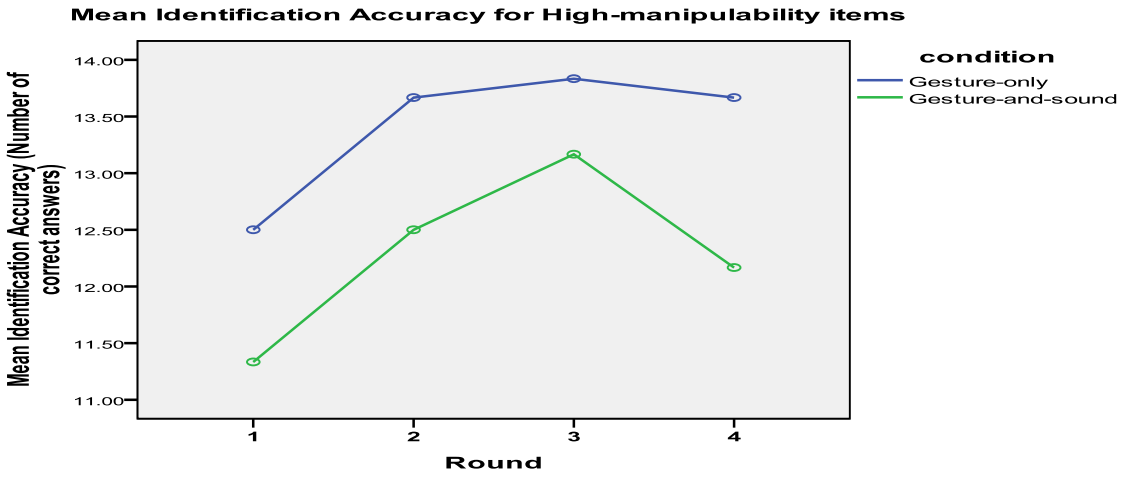
(I) condition	(J) condition	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
Gesture Only	Gesture and Sound	1.688*	.429	.003	.731	2.644

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.





Appendix D: Gesture Length / Complexity (Time variable)

By-subject Analysis

Mauchly's Test of Sphericity (By-subject analysis)

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
round	.063	29.600	5	.000
manipulability	1.000	.000	0	.
round * manipulability	.146	20.614	5	.001

Tests of Within-Subjects Effects (By-subject analysis)

Source		df	F	Sig.
round	Sphericity Assumed	3	65.390	.000
	Huynh-Feldt	1.468	65.390	.000
round * condition	Sphericity Assumed	3	1.322	.282
	Huynh-Feldt	1.468	1.322	.282
Error(round)	Sphericity Assumed	36		
	Huynh-Feldt	17.621		
manipulability	Sphericity Assumed	1	3.904	.072
	Huynh-Feldt	1.000	3.904	.072
manipulability * condition	Sphericity Assumed	1	.083	.778
	Huynh-Feldt	1.000	.083	.778
Error(manipulability)	Sphericity Assumed	12		
	Huynh-Feldt	12.000		
round * manipulability	Sphericity Assumed	3	2.541	.072
	Huynh-Feldt	1.966	2.541	.101
round * manipulability * condition	Sphericity Assumed	3	.558	.646
	Huynh-Feldt	1.966	.558	.577
Error(round*manipulability)	Sphericity Assumed	36		
	Huynh-Feldt	23.586		

Tests of Between-Subjects Effects (Condition) (By-subject analysis)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	225182.893	1	225182.893	453.957	.000
condition	4810.321	1	4810.321	9.697	.009
Error	5952.536	12	496.045		

Pairwise Comparisons (Condition) (By-subject analysis)

(I) condition	(J) condition	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
gesture only	gesture and sound	-13.107*	4.209	.009	-22.278	-3.936

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Univariate Tests (Condition) (By-subject analysis)

	Sum of Squares	df	Mean Square	F	Sig.
Contrast	601.290	1	601.290	9.697	.009
Error	744.067	12	62.006		

Pairwise Comparisons (Manipulability) (By-subject analysis)

(I) manipulability	(J) manipulability	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
Low Manipulability	High Manipulability	3.429	1.735	.072	-.352	7.209

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

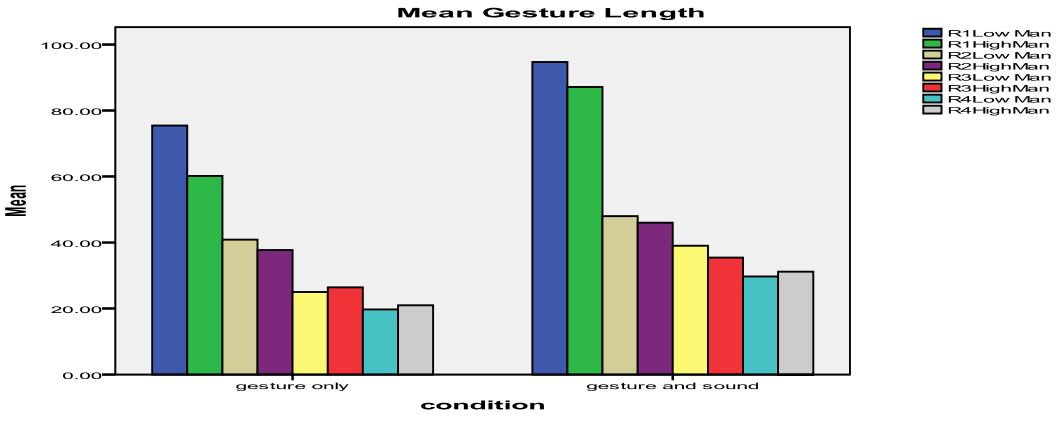
Pairwise Comparisons (Round) (By-subject analysis)

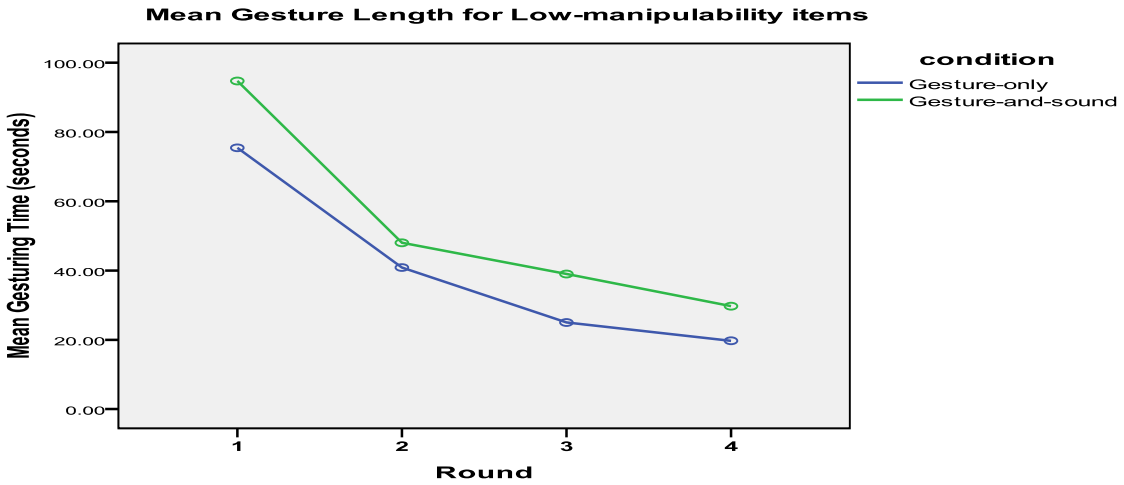
(I) round	(J) round	Mean Difference (I-J)	Std. Error	Sig.[a]
1	2	36.214*	5.497	.000
	3	47.893*	5.562	.000
	4	53.964*	6.002	.000
2	1	-36.214*	5.497	.000
	3	11.679*	1.902	.000
	4	17.750*	2.113	.000
3	1	-47.893*	5.562	.000
	2	-11.679*	1.902	.000
	4	6.071*	1.370	.005
4	1	-53.964*	6.002	.000
	2	-17.750*	2.113	.000
	3	-6.071*	1.370	.005

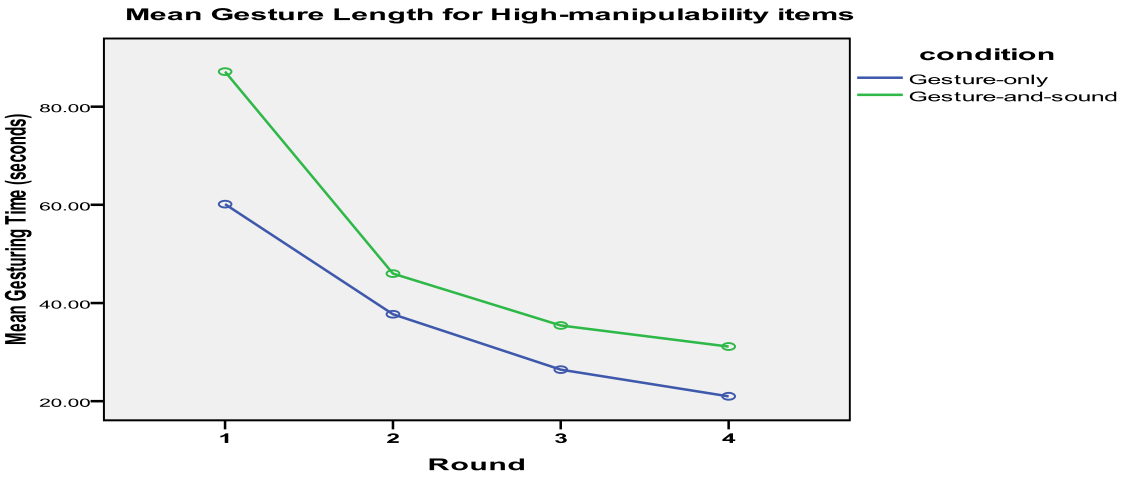
Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.







By-item Analysis

Mauchly's Test of Sphericity (By-item analysis)

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
Round	.184	14.767	5	.012
Condition	1.000	.000	0	.
Round * Condition	.118	18.630	5	.002

Mauchly's Test of Sphericity (By-item analysis)

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
Round	.184	14.767	5	.012
Condition	1.000	.000	0	.
Round * Condition	.118	18.630	5	.002

Tests of Within-Subjects Effects (By-item analysis)

Source		df	F	Sig.
Round	Sphericity Assumed	3	93.358	.000
	Huynh-Feldt	1.879	93.358	.000
Round * Manipulability	Sphericity Assumed	3	1.210	.323
	Huynh-Feldt	1.879	1.210	.318
Error(Round)	Sphericity Assumed	30		
	Huynh-Feldt	18.795		
Condition	Sphericity Assumed	1	16.749	.002
	Huynh-Feldt	1.000	16.749	.002
Condition * Manipulability	Sphericity Assumed	1	.012	.914
	Huynh-Feldt	1.000	.012	.914
Error(Condition)	Sphericity Assumed	10		
	Huynh-Feldt	10.000		
Round * Condition	Sphericity Assumed	3	2.634	.068
	Huynh-Feldt	1.655	2.634	.109
Round * Condition * Manipulability	Sphericity Assumed	3	.353	.787
	Huynh-Feldt	1.655	.353	.668
Error(Round*Condition)	Sphericity Assumed	30		
	Huynh-Feldt	16.549		

Pairwise Comparisons (Manipulability) (By-item analysis)

(I) Manipulability	(J) Manipulability	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
Low	High	3.833	6.469	.567	-10.580	18.247

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

Univariate Tests (Manipulability) (By-item analysis)

	Sum of Squares	df	Mean Square	F	Sig.
Contrast	44.083	1	44.083	.351	.567
Error	1255.380	10	125.538		

Pairwise Comparisons (Round) (By-item analysis)

(I) Round	(J) Round	Mean Difference (I-J)	Std. Error	Sig.[a]
1	2	42.333*	4.872	.000
	3	55.750*	5.243	.000
	4	62.833*	5.819	.000
2	1	-42.333*	4.872	.000
	3	13.417*	2.029	.000
	4	20.500*	2.795	.000
3	1	-55.750*	5.243	.000
	2	-13.417*	2.029	.000
	4	7.083	2.174	.052
4	1	-62.833*	5.819	.000
	2	-20.500*	2.795	.000
	3	-7.083	2.174	.052

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Pairwise Comparisons (Condition) (By-item analysis)

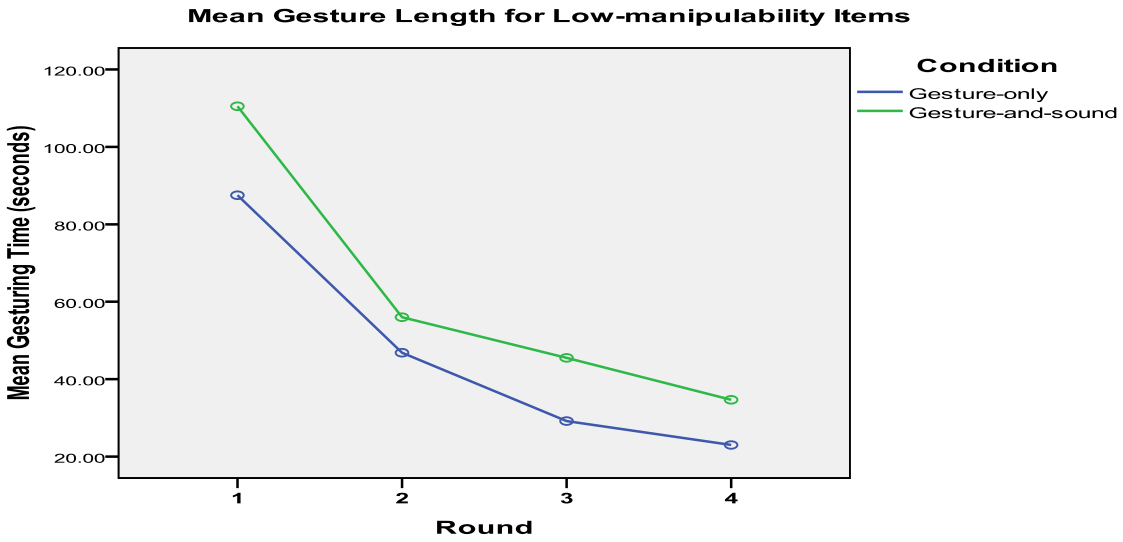
(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]

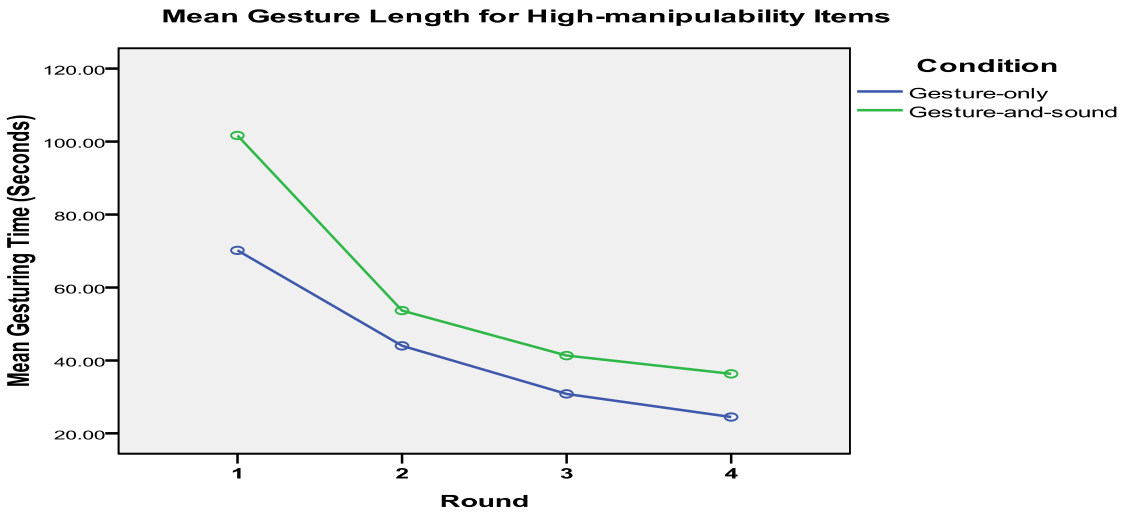
					Lower Bound	Upper Bound
Gesture-and-sound	Gesture-only	15.458*	3.777	.002	7.042	23.874

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.





Appendix E: R1-4 & R5 Identification Accuracy Correlation (Gesture-and-sound Condition)

Correlations

		R1to4	R5
R1to4	Pearson Correlation	1	.307
	Sig. (2-tailed)		.504
	N	7	7

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.307[a]	.094	-.087	2.85552

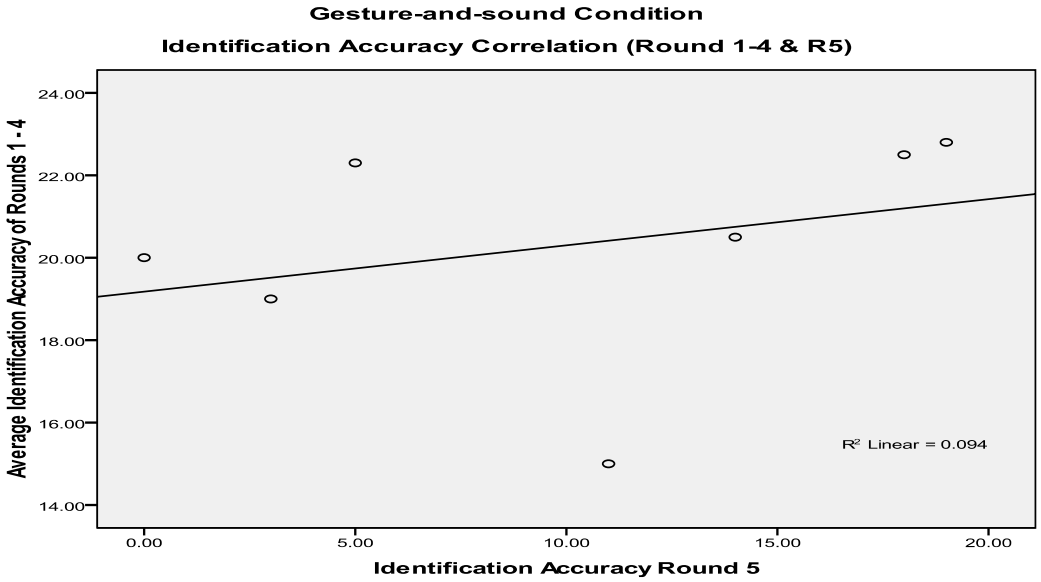
a. Predictors: (Constant), R5

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.230	1	4.230	.519	.504[a]
	Residual	40.770	5	8.154		
	Total	45.000	6			

a. Predictors: (Constant), R5

b. Dependent Variable: R1to4


Gesture-and-sound Condition
Identification Accuracy Correlation (Round 1-4 & R5)
Average Identification Accuracy of Rounds 1 - 4
24.00
22.00
20.00
18.00
16.00
14.00
0.00
5.00
10.00
15.00
20.00
Identification Accuracy Round 5
R² Linear = 0.094